Master's Thesis

Identification of various user group in university counseling scene using machine learning algorithm

Kwanglo Lee

Department of Biomedical Engineering

(Human Factors Engineering)

Ulsan National Institute of Science and Technology

2021

Identification of various user group in university counseling scene using machine learning algorithm

Kwanglo Lee

Department of Biomedical Engineering

(Human Factors Engineering)

Ulsan National Institute of Science and Technology

Identification of various user group in university counseling scene using machine learning algorithm

A thesis/dissertation submitted to
Ulsan National Institute of Science and Technology
in partial fulfillment of the
requirements for the degree of
Master of Science

Kwanglo Lee

12/23/2020

Approved by

Advisor
Dooyoung Jung

Identification of various user group in university counseling scene using machine learning algorithm

Kwanglo Lee

This certifies that the thesis/dissertation of Kwanglo Lee is approved.

12/23/2020

Signature

Advisor: Dooyoung Jung

Signature

Sung Phil Kim: Thesis Committee
Member #1

Signature

Chiehyeon Lim: Thesis Committee
Member #2

ABSTRACT

There are increasing need for improving university students' mental health. Still, the quantitative and qualitative abilities of the university counseling center are insufficient to handle the increasing need and severity. In addition, it is difficult for individual counseling centers to solve such problems due to lack of numbers in counselors and low budget. Therefore, to solve current problems by increasing the service efficiency of the university counseling center service, identification of various user groups was done using machine learning algorithm based on initial stage data of counseling service. To be specific, user service, additional clinical and latent group was identified to reduce counselor's burden at initial stage of service and provide reference for clinical decision and future service planning.

This study utilized data acquired from UNIST healthcare center and analyzed in two major steps. First, service (counseling or clinical treatment with drug use) and clinical (suicidal-risk and potential dropout) group classification was done using supervised learning algorithms and identified important feature in classifying each group. Then, the latent groups reflecting the detailed characteristics of users was analyzed by using latent class analysis. Laten user group identification detected 5 different latent groups (lower risk, lower/moderate risk, moderate risk, higher risk with sleep issue, sleep problem group) and their distinctive characteristics.

Current study successfully detected meaningful reference for university counseling service using data focused on the initial stage of the service. Analyzing service effectiveness and using it as reference for converting counseling service into clinical treatment according to current study results will increase overall service effectiveness provided to individual users. The result of suicidal-risk group classification identified similar features from prior researches without using additional screening tools. Interestingly, dropout group classification results identified features that were not found in prior research which can be used in future service planning to prevent user dropout during the service. After the classifications of various user groups were conducted, improving the result of ensemble modeling using stacking classifier was done to achieve higher performance in type 2 error of classification results. Latent group identification found sub-groups that can be applicable to existing counseling services and possible customized clinical approaches can be provided to individual latent groups.

Further studies including 1) improving machine learning algorithm performance by developing data collection methods that reflect user characteristics 2) better, service effectivity analysis using overall service records and 3) applying studied researches to other university counseling centers will also contribute to reducing individual counselor's burden and enhancing university counseling center service effectiveness.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

Being a college student is the first stage of entering society as an adult and a transitional period going through problems students did not experience before. In this period, college students experience various social anxiety factors such as economic difficulties, determining careers and psychological difficulties such as interpersonal relationships, adaptation to new environment, and physical-sexual violence.(이혜선 et al., 2012)

Notably, the mental health problems of college students are directly related to the mental health problems of young people in the 20s. The domestic university entrance rate was 70% and ratio of college undergraduate and graduate student is 54.2% among all 20s in 2018. In addition, the number of patients with mental health disease in 20s increased the most among all age groups for last five years in 2018, and the most frequent disease was identified as 'depression'. (건강보험심사평가원, 2018). In addition, the ratio of experiencing severe depressive episode in the 20s increased from 9.3% to 14.9% between 2012 to 2015.(노윤신 & 이성은, 2019)

Accordingly, the necessity and responsibility of the university-affiliated counseling center, which is responsible for the mental health of university members including college students and graduate students, is increasing. In addition, the mental health problems that university members report are becoming more serious and diversified, and university counseling centers are required to have a higher level of competence.(Flynn & Heitzmann, 2008; 김은하, 전소연, & 김다예, 2016) Particularly, the university counseling center has its advantages in low psychological and economic barrier to the service, easy tracking and management of users, and possible early detection of mental health problems. For this reason, university counseling center is a key area of managing of college mental health.

However, the mental health management system of domestic universities is insufficient in both qualitative and quantitative way. For many domestic university counseling centers, the ratio of part-time counselors is high to cover the lack of numbers in professional counselors. As a result of the survey of seven universities of various sizes in Korea, the university counseling center consisted of two or three full-time counselors and many other of part-time counselors. In addition, there are significant differences in number of students per counselors compared to US universities. Domestic university counseling centers hold about 2,000 students per counselor while US universities' mean students per counselor is about 700. As a result, newly registered user takes about one month or two to receive counseling service. In addition, the proportion of high-risk students like suicidal-risk group who visit university counseling centers is increasing from 1.5% to 10%.(노윤신 & 이성은, 2019)

However, there was no counseling centers where established the system for immediate screening and intervention for students who needed medication among high-risk groups. Most service protocols go through similar process except for outstanding suicidal-risk group. Therefore, mental health risk factors can be detected at reception interview or after the service begin. Moreover, the reception interviews in many university counseling centers are often conducted by intern counselors due to lack of numbers in professionals. Due to this reasons, sometimes users' symptoms and problems are mis-diagnosed, and some problems that could been resolved earlier cannot be handled properly or even get worse.(이영은, 차영은, & 민경화, 2013) However, many counselors report that it is difficult to expand the capacity of counseling centers to handle the increasing demand due to lack of understanding at the university headquarters about the importance of the counseling center.(노윤신 & 이성은, 2019)

To solve such problems, current study conducted various user group identification using machine learning algorithm to improve service efficiency of university counseling centers. Specifically, this study aims to conduct user group classification based on counseling service application data using machine learning algorithm and provide reference for clinical decision-making and future service planning. We classified users into counseling or clinical treatment group based on the service application form submitted by users and detect other potential groups.

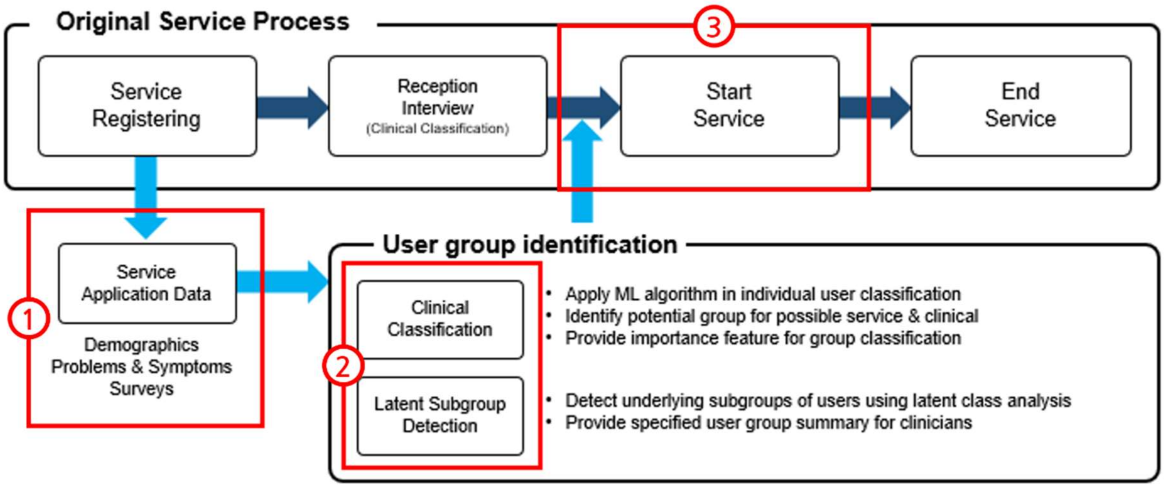


Figure 1. Application example of current research result to existing counseling service process

Figure 1 is the application example of current research results to existing counseling service in university counseling center. By applying current results, 1) data collection without additional burden to individual counselors is possible by using data that is already collected at initial stage of service. And 2) the user service/clinical group using machine learning algorithm and the latent user group identification can reduce the burden of clinical decision-making at the beginning of the service and obtain information about the characteristics and demographics of the users visiting the university counseling center. Lastly, 3) the information about user can be provided to clinicians at the early stage of service and more suitable treatment for each user and active early intervention can be done which enhance overall counseling service effectiveness.

II. RELATED WORKS

Current chapter will introduce related prior researches and their strength and weakness while discussing about focus of current study over them.

Machine learning models are gradually expanding their use in the field of mental health. In particular, machine learning models are effectively applied to the areas including diagnosis, prevention and treatment. To be specific, diagnosing and predicting mental health issue is one of the most common researches conducted. Supervised learning models are used to predict future mental disorder and detect important features in prediction which were previously difficult to identify using given data.(Cho, Yim, Choi, Ko, & Lee, 2019)

For example, various machine learning models can be applied to predict future mental health problems. Study on predicting future mental health problems was conducted. In the prior study, the possible development of mental health problems was tested in adolescence based on the mental health indicators of childhood using supervised learning algorithms. Supervised learning models such as Logistic Regression, Random Forest, XGBoost, Neural Network, Support Vector Machine were applied and ROC AUC was used as the performance indicator. The study suggested that machine learning models can have a sufficient effect on mental health problems compared to simple Logistic Regression. In addition, the study derived feature importance using tree-base models to test which features affect the most in prediction.(Tate et al., 2020)

However, it focused on general population rather than specific target group. In addition, overall competency of machine learning models was tested which neglected implications to actual service. Current study aims to take a step further in testing specific target group, college students visiting college counseling center. Moreover, current research focus on the initial stage of service to identify information that can be delivered to the counselors using data collected at that stage.

Unsupervised learning is another form of machine learning technique which is not commonly used as supervised learning in mental health issues. It is because many studies try to predict psychiatric problems based on existing diagnosis criteria. Still, unsupervised learning has its advantage in the possibility of detecting unknown insights which cannot be found using existing labels. It is especially effective in identifying groups and characteristics that was not able to find using pre-determined target values.(Cho et al., 2019)

Latent class analysis is one of the supervised clustering algorithms. It is originally used for user group segmentation. A prior study utilizing latent class analysis applied it to detect health-risk behavior pattern of college students. It attempted to find latent and comprehensive structure of user group which was hard to detect using existing screening tools. As a result, 3 different latent groups(typical, high-risk, moderate healthy) were detected. In addition, the features related to each latent groups were analyzed to suggest which factors to focus in future service.(Kwan, Arbour-Nicitopoulos, Duku, & Faulkner, 2016)

It has the advantage of discussing the specific group of college students and various groups and characteristics to be dealt with in future services. However, the study utilized data recorded in the form of self-report on symptoms, not the actual screening method. In current study, data collected from the initial stage of the actual service was utilized. It is expected that this will reduce the burden on counselors and apply the information provided more efficiently to the service by utilizing the existing diagnosis criteria.

III. METHOD

This chapter describes the data used in the project and the preprocessing procedures, and then the detailed explanation of the research progress.

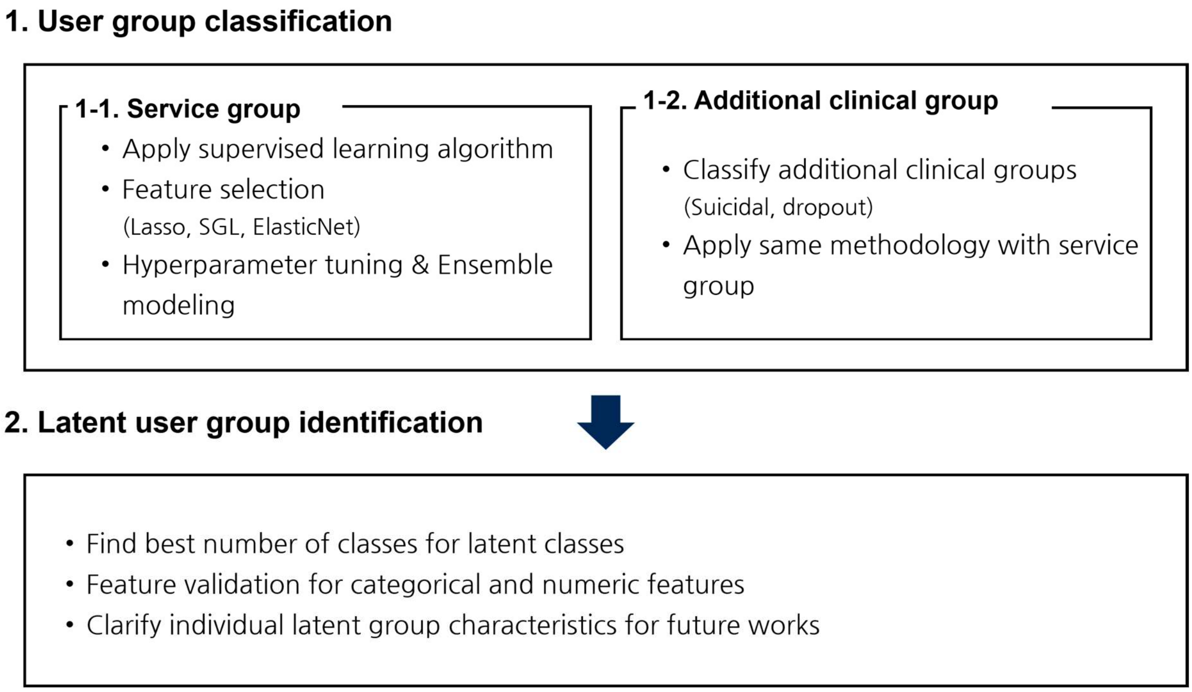


Figure 2. Overall flow of current research

Figure 2 is the overall flow of the current research. Current study consist of two main parts, user group classification and latent user group identification. User group classification utilize supervised learning algorithms to classify two sub categories. First sub category of user group classification is user service group. Service group is the most basic classification used in UNIST healthcare center determining whether to provide counseling service or need of clinical treatment with the psychiatrist using drugs. Reducing counselor's burden in classifying users at early stage of service can be achieved by providing predicted user service group. In addition, it is possible to continuously manage the severity change of the user and the service effectiveness by detecting important characteristics in distinguishing counseling and clinical treatment.

Second sub category for user group classification is additional user clinical group classification. After service group classification, feedback session with UNIST healthcare center counselors was held. In the feedback session, discussion about further steps of the project and insights related to healthcare service was provided by counselors. Additional clinical group such as suicidal-risk and dropout-risk group was suggested as possible and critical groups in managing actual counseling service. Suicidal-risk group is one of the most urgent group that needs to be handled with immediate intervention and continuous monitoring. Thus, early identification and monitoring related features from the beginning of the service is required even though certain user is not considered as high-risk in suicidal risk. Such approach can detect suicidal tendency in early stage and prevent possible suicidal events. Dropout refers to sudden and unexpected stopping of the service without mutual agreements with the counselor. Dropout can lead to unexpected problems in that constant monitoring of the user's status cannot be done properly. Therefore, it is necessary to detect potential risk factors in advance and monitor related indicators to prevent dropout event.

Second main part latent user group identification was conducted using latent class analysis(LCA) which is commonly used unsupervised learning algorithm in customer segmentation. The aim of latent user group identification was to cluster the various type of users visiting university counseling center. As for service group, there are various severity in problems and symptoms within and between counseling and clinical treatment group and such variety cannot fully represent individual user's needs. Suicidal and dropout group represent important clinical group that requires constant monitoring. But the such group stands out as an outlier and cannot be used as representative feature for overall users. Thus, by conducting latent user group identification, acquiring better insight towards various user group is needed apart from existing clinical labels. Such results can identify specific user's needs more intuitively and provide group characteristic and current severity of each user within latent group to enhance user's understanding about current individual status.

3.1 Data Introduction

3.1.1 Structure of service application form and feature characteristics

Current study was conducted based on total of 343 clinical records of users who visited UNIST healthcare center between March 13, 2018 to December 27, 2019. Clinical records included from service application(registration) data and initial mental health screening results(surveys), 8, 12th week follow-up screenings and evaluation of user at the end of the service. In this study, only the initial application data and some related service records were used according to the purpose of the study. The following are the descriptions of the items and questionnaires from the collected data used in the study.

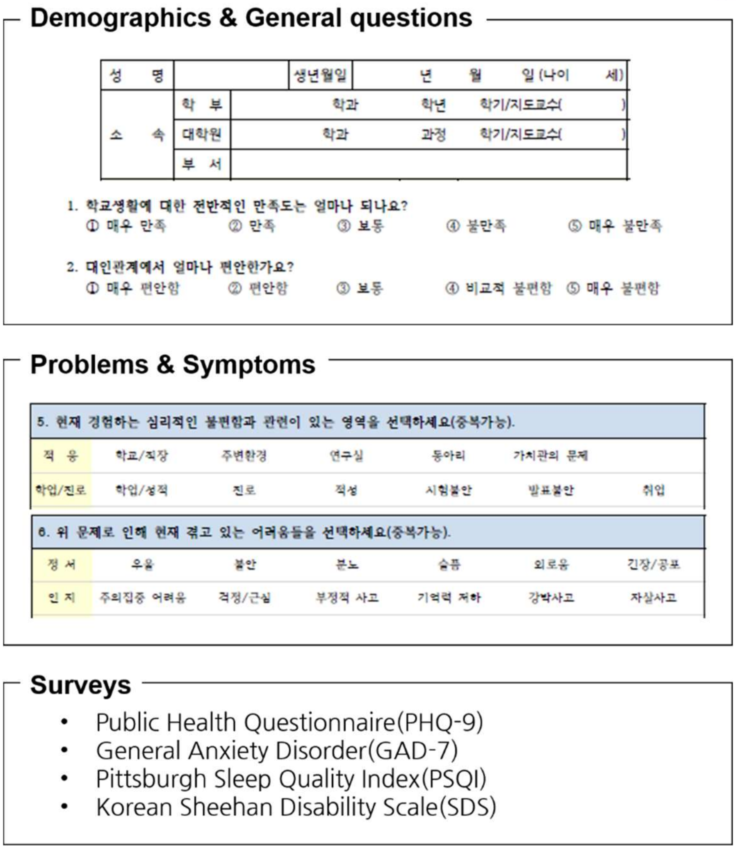


Figure 3. Structure of UNIST healthcare center service application form

Demographics & General questions

The first part of the data, demographics and general questions, included many personal information, and went through the screening of counselors in advance. Through the screening procedure, data was transformed into generalized information such as age, gender, housing type, and affiliation. General questions include satisfaction to school, relationship, self and also contains self-reports about cause(problem) and symptoms related to difficulty users are currently suffering. In addition, other questions like the experience of previous help from professionals, family history of mental disorder and suicidal attempts are also included. Total application form is available at appendix. After demographics and general questions, screening surveys are present to measure user's current mental health status.

Public Health Questionnaire(PHQ-9)

To measure the severity of depressive symptoms, Korean translated version of public health questionnaire(PHQ-9) was used.(Kroenke, Spitzer, & Williams, 2001; 안제용, 서은란, 임경희, 신재현, & 김정범, 2013) PHQ-9 evaluate 9 major depressive symptoms according to diagnosis criteria from Diagnostic and Statistical Manual of Mental Disorders 4th edition(DSM-4). Each question is rated in 4-point scale(0~3) about suffering depressive symptoms within 2 weeks. Total PHQ score range from

0 to 27 and PHQ categories are divided into 0~4(Not depressed), 5~9(mild depression), 10~19(moderate depression), 20~(severe depression).

General Anxiety Disorder(GAD-7)

Korean version of general anxiety disorder scale was used to identify anxiety symptoms. GAD-7 was developed for diagnosing general anxiety disorder and evaluating symptom severities.(Spitzer, Kroenke, Williams, & Löwe, 2006) Questions are marked in 4-point scale(0~3) about suffering anxiety symptoms within 2 weeks. Total range of GAD score is 0~21 and GAD category is divided into 5~9(mild anxiety), 10~14(moderate anxiety) and 15~(severe anxiety). Validation of Korean translated version was done with epilepsy patients.(Seo et al., 2014)

Pittsburgh Sleep Quality Index(PSQI)

Pittsburgh Sleep Quality Index(PSQI) consist of 18 different questions measuring 7 specific components: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbance, use of sleeping medication, daytime dysfunction.(Buysse, Reynolds III, Monk, Berman, & Kupfer, 1989) Each question has 4-point scale(0~3) and each component is calculated with mixture of different question results and transformed into also 4-point(0~3) scale. Result of each component refers to 0(no sleep problem), 1(mild sleep problem), 2(moderate sleep problem) and 3(severe sleep problem). Total PSQI score range from 0~21 by adding all scores of individual components. If total PSQI score is less than 5, it is considered as good sleeper and over 5 refers to poor sleeper.

Korean version of Sheehan Disability Scale (SDS)

Disability scale was originally developed for measuring daily functional disability of social anxiety patients.(Sheehan, 1983) This scale was proved to be also effective in evaluating functional disability of other anxiety and depression patients.(Leon, Olfson, Portera, Farber, & Sheehan, 1997) Functional disability scale measures difficulty in occupational/academic, social, family due to functional disability using self-report method. Validation of Korean translated version was conducted with psychiatric patients.(박준영 & 김지혜, 2010)

3.2 Research methods

3.2.1 Data pre-processing

After receiving pre-screened data from UNIST healthcare center, data pre-processing was done to convert service records into format that can be used in research.

Delete unnecessary data(Data selection)

First step of pre-processing was choosing data columns to include at the analysis. Since the data was originally made for service record managements, some unnecessary data was deleted. Considering the focus of current study, 8-week and 12-week follow-up records were deleted since it was unable to acquire at the initial stage of service. Variables from the final part of the service like 'final treatment provided' and 'service termination type' was included as the target value for the user group classification.

Data adjustment and transform

Raw data was delivered in excel format and transformed into format capable of use in R and Python. In addition, since the data was hand-recorded to excel spread sheet, there were some misinput in values. For example, some columns with binary input(0,1) had 2 as its components and some multiple input separator was marked as '.' not ','. There were also some missing values which were treated differently according to each column characteristic. Columns related to problems or symptoms assumed missing value as not suffering such issues and median value was used for numeric values to minimize the effect to total column. For cases answering to daily habits related multiple questions as depends on daily condition, such answers were changed to median value and new column named 'question_random' was added to record such irregular habit was present. Variables with multiple answers in single columns, binary encoding was used to separate each answers to individual columns with 0, 1. In addition, some responses with short sentences where summarized to 0(None), 1(Present). For example, response to 'family history of mental disorder' contained multiple diagnosis and explanations. Other survey questionnaires which require ratings were proceeded with their reference papers to calculate survey score.

3.2.2 User group classification

1) User service group classification

The first part of the user service group classification aims to classify two main division of service, counseling and clinical treatment with psychiatrist, in UNIST healthcare center. The target value for the classification was based on service record 'treatment final' representing which service was provided to certain user at the final service session. The classification results were divided into counseling - 0 and clinical treatment with psychiatrist – 1. Counseling refers to receiving just counseling service and clinical treatment group received both counseling and medical treatment(drug) under supervision of professional psychiatrist.

Multiple supervised learning algorithms were applied for the user service group classification. The algorithms used for classifications were: Logistic Regression(LR), K-Neighbors(KNN), Linear

Discriminant Analysis(LDA), Support Vector Machine(SVM), Multi Layer Perceptron(MLP), Random Forest(RF), Extra Trees(ExT), AdaBoost(Ada), XGBoost(XGB). All supervised models used in current study were functions provided from Python scikit-learn library.(Pedregosa et al., 2011) To reproduce the classification results, train-test ratio was set to 8:2 and random state was fixed to 2020. The model performance indicator was receiver operating characteristic area under curve(ROC AUC) for applying different threshold for diverse severity and symptoms. Feature selection methods for identifying important features in classifications were Lasso, Sparse-group Lasso(SGL) and ElasticNet. Lasso and ElasticNet was from Python scikit-learn library and SGL was from R 'SGL' package.(Simon, Friedman, Hastie, Tibshirani, & Simon, 2018)

Lastly, for model performance optimization, hyperparameter tuning and ensemble modeling was conducted. Hyperparameter was done by grid search and soft voting and stacking ensemble modeling was used for ensemble modeling. The default model for stacking classifier was LR and other models were tested at model performance improvement for reducing type 2 error. To adjust the difference in ensemble modeling results at every iteration, 100 repetition was done per single model and its mean performance was used for performance indicator. Both grid search and ensemble modeling used functions provided by Python scikit-learn library. After ensemble modeling, confusion matrix of each model result was compared and improvements in ensemble modeling to minimize type 2 error was done.

2) Additional clinical group classification

Additional clinical group classification started from target value selection. For suicidal-risk group, there was no direct service record for classifying suicidal-risk group. Thus, current study selected 'Q14 suicidal attempt' column as target value which indicate previous attempts of suicide. Dropout-risk group used service record 'service termination' which recorded how the service ended labeled as early termination(without contact). Other classification methodology was identical to the user service group classification.

3.2.3 Latent user group identification

After the user group classification, unsupervised learning algorithm was applied to identify specific latent user groups that cannot be detected with existing data labels. Selected unsupervised learning algorithm was latent user analysis(LCA) which is widely used in customer group segmentation. To find the most appropriate latent group number, LCA using different class number from 2 to 7 was conducted and 200 iterations were done per class to find better model in clustering. Model selection criteria was Bayesian information criterion(BIC). After selecting the best number of classes for the analysis, comparison analysis according to variable type(numeric or categorical) was conducted for model validation. For numeric variables, partial η^2 measuring the effect in group classification was used and Cramer's V for categorical variables. Latent class analysis used 'depmixS4' package from R. (Visser & Speekenbrink, 2016) 'depmixS4' use expectation-maximization(EM) algorithm which is one of the Gaussian mixture model(GMM) to identify latent groups. After validation of each variable type, analysis of individual group characteristics using validated features were done. Current study used Python 3.6.9 and R 3.6.3 for analysis.

IV. RESULTS

4.1 User service group classification

The service group classification was conducted using the preprocessed data. Based on the results, the clinical feedback on the project was discussed, and after several adjustment steps, the service group classification using same process was conducted with modified data. Feature selection methods were applied after the classification to select importance features in classifying each service group. After feature selection, hyperparameter tuning and ensemble modeling was done for model performance optimization. Then, improvements in ensemble modeling were conducted using ROC AUC and confusion matrix results.

4.1.1 Service group classification result using all data

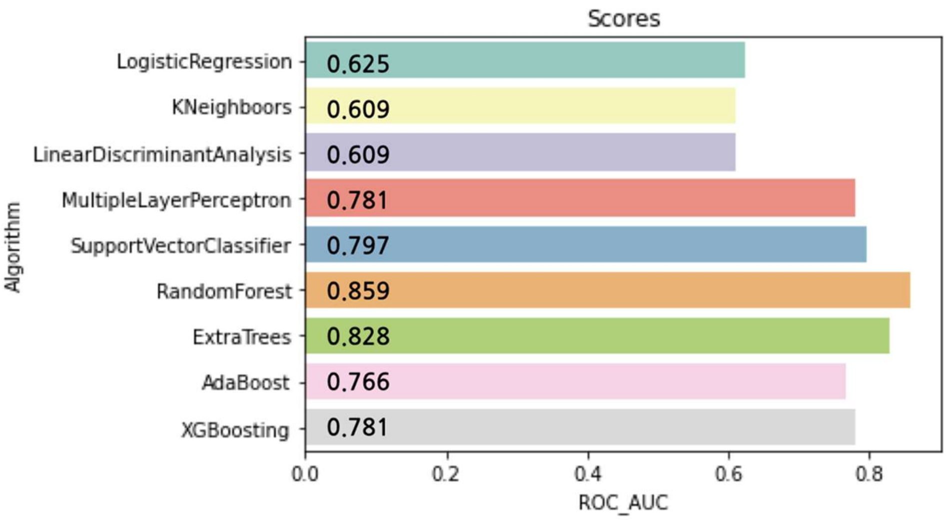


Figure 4. Cross validation result using total pre-processed data

First analysis was the user service group classification using the raw preprocessed data. As in figure 4, cross validation results indicated that bagging ensemble models like RF and ExT performed slightly better compared to other models. All other models showed similar performance apart from relatively simpler models like LR, KNN and LDA.

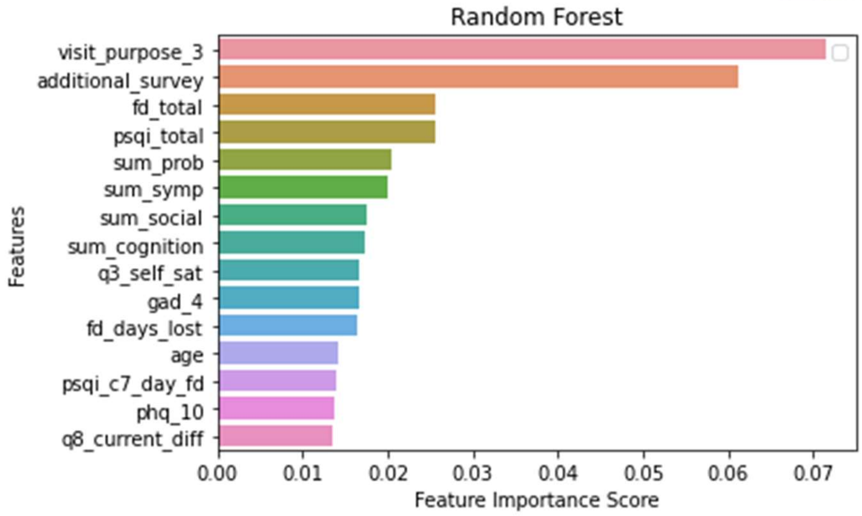


Figure 5. Feature selection result in Random Forest

In addition to results in figure 4, the feature importance in classifications using tree-based models were tested. As a result, 'visit_purpose_3' indicating the purpose of visiting counseling center was clinical treatment and 'additional_survey' meaning presence of record of 8, 12-week follow up surveys were selected as important features in classifying counseling and clinical treatment group. Since the present of additional follow-up survey was not able to be identified at initial stage of the service, 'additional_survey' was deleted from data. Afterwards, the counselors' feedback was held in UNIST healthcare center to evaluate result of current study and receive insights about healthcare services. During counselor feedback, some columns that were misunderstood were fixed and discussed about the need of additional clinical group classifications. In addition, clinical discussion about practical application of current study was done.

4.1.2 Service group classification result after counselor feedback

Based on the classification results using the raw pre-processed data, feedback was given from clinicians of the healthcare center. As a result, the purpose of the visit('visit_purpose') is often changed after the reception interview, and the previous classification results showed the classification bias due to explicitly wishing for counseling/clinical treatment. In addition, the results of the discussion that variables such as 're-application', which means that the user registered again after the service termination, can distort the seriousness of the users were excluded.

1) Feature selection results

Feature selection was done by using Lasso with 10-fold cross validation, Sparse-group Lasso and ElasticNet. As a result, variables in table 1 was selected.

Variable	SGL	LassoCV	ElasticNet
Q3 Self satisfaction	■	■	■
Q8 Current difficulty	■	■	■
Q9 Experience of past professional treatment		■	■
Q14 Suicidal attemp	■	■	■
PHQ - Difficulty due to depressive feelings	■	■	■
PHQ - Tasteless or overeating		■	
PHQ - Suicidal/self-injurying thoughts	■		
GAD - Difficulty staying still	■	■	■
GAD - Anxiety symptom	■	■	■
GAD - Difficulty being comfortable	■	■	
PSQI score total	■	■	■
PSQI category	■	■	■
PSQI - Funtional disability due to sleep issue	■	■	■
PSQI - Taking sleep-related drugs	■	■	■
PSQI - Other sleep problems		■	
Physiological symptoms (Total)	■	■	■
Panic symptom	■	■	■
Physical symptom	■	■	■
Ataraxia symptom	■	■	
Obsessive and compulsive action symptom	■	■	
Suicidal thought symptom		■	
Religious problem	■	■	■
Sibling problem		■	■

Table 1. Common feature selection results in user service group classification

'Variable' in table 1 means the name of each selected variable, and each row filled with yellow in SGL, Lasso, and ElasticNet column means that a specific variable is selected. As a result of feature selection, various features were selected in general satisfaction, suicide problem, depression and anxiety, sleep problem, various symptoms and problems areas. The selected features had many variables in common and included features from various areas. This means that consideration is needed in various areas, not just one or two areas, to distinguish between counseling and clinical treatment. This is in line with the prediction that there will be various seriousness and user demand inside the service group.

2) Classification results using cross validation

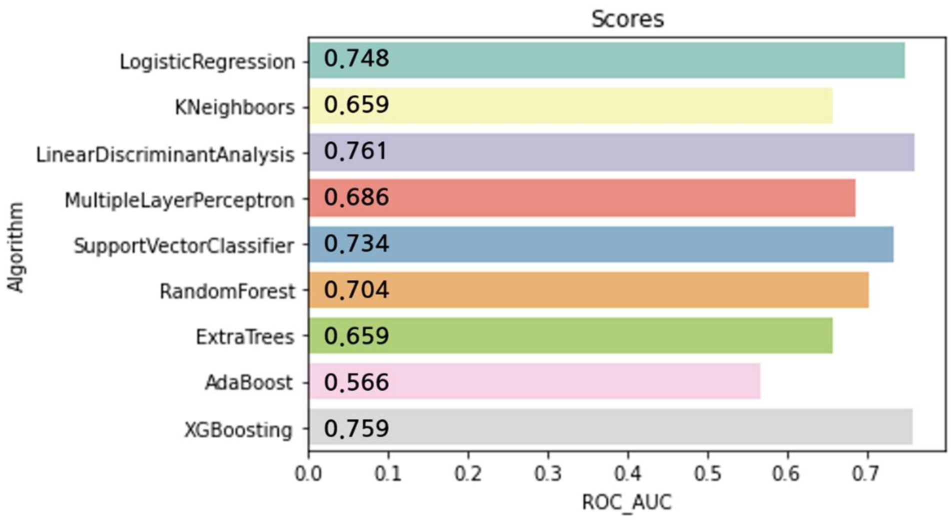


Figure 6. User service group classification results using SGL features

Based on the features selected through the feature selection methods, the same analysis was conducted using machine learning algorithm. As a result, the features selected using SGL and ElasticNet showed better performance than Lasso. Although there were some differences according to individual model, SGL and ElasticNet showed similar performance.

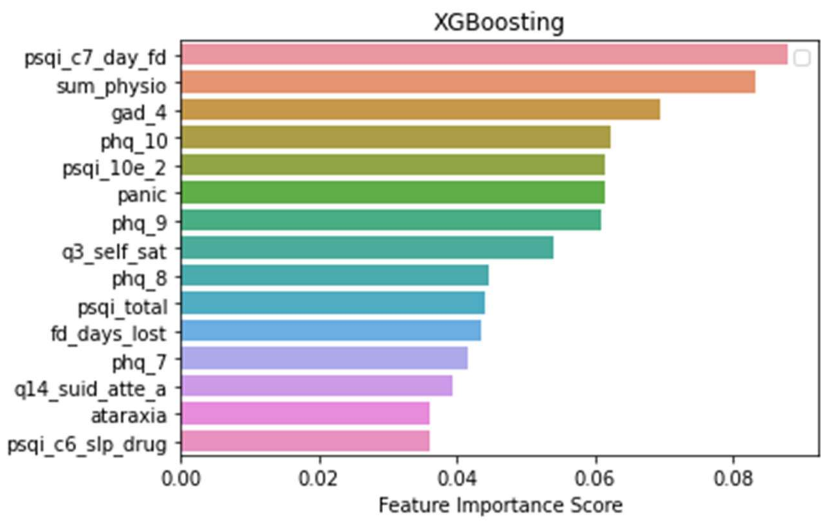


Figure 7. Feature importance from XGBoost in user service group classification

In Figure 7, by reviewing the feature importance from XGBoost, features like "Daytime functional disability due to sleep(psqi_c7_day_fd)", "Suffering from physiological symptoms(sum_physio)", "current difficulty due to depressive symptoms(phq_10)", "presence of other sleep problems (psqi_10e_2)" were selected as the core features classifying between counseling and clinical treatment.

3) Hyperparameter Tuning

Model Name	CV Score (ROC AUC)	Best Score (ROC AUC)
Logistic Regression	0.748	0.751
K-Nearest Neighbors	0.659	0.738
Linear Discriminant Analysis	0.761	0.730
Support Vector Machine	0.686	**0.766**
Multi Layer Perceptron	0.734	**0.774**
Random Forest	0.704	0.750
Extra Trees	0.659	**0.768**
AdaBoost	0.566	0.717
XGBoost	0.759	0.721

Table 2. Hyperparameter tuning results using features using ElasticNet in user service group classification

After feature selection, hyperparameter tuning was conducted for model optimization. Hyperparameters used for each model is in table A1 at the appendix. As a result of hyperparameter tuning in table 2, model using features selected by ElasticNet resulted best performance. And SVM, MLP, ExT showed best results and especially enhanced in result of hyperparameter tuning. This may be due to individual model characteristics like dealing with high-dimensional and non-linear features(SVM, MLP) and evaluating features in more broader view(ExT). Bagging models over boosting results show that being robust to overfitting also showed better results. Detailed selected hyperparameters are in table A2.

4) Ensemble Modeling

Classifier	Feature Selecting Method	Mean of 100 iterations
Voting Classifier	Lasso	0.69483
	Sparse-group Lasso	**0.73733**
	ElasticNet	0.73633
Stacking Classifier	Lasso	0.6625
	Sparse-group Lasso	**0.73167**
	ElasticNet	0.70933

Table 3. Ensemble modeling result of user service group classification

Table 3 is result of ensemble modeling and features from SGL showed best results. In voting classifier, SGL and ElasticNet did not show much difference, but SGL showed better performance in stacking classifier. In addition, apart from ROC AUC, it is important to reduce the false negative, type2 error, in this study. Therefore, the confusion matrix for each ensemble modeling was compared to verify the results.

Voting	**Predicted Counseling**	**Predicted Clinical treatment**
Actual Counseling	28	9
Actual Clinical treatment	6	17

Stacking	**Predicted Counseling**	**Predicted Clinical treatment**
Actual Counseling	27	10
Actual Clinical treatment	6	17

Table 4. Confusion of matrix service group classification using SGL features

In comparing the result of the confusion matrix, ensemble modeling using features from SGL showed the lowest type 2 error in classifying clinical treatment over counseling.

4.2 Additional clinical group classification

After the service group classification, same procedure was applied to suicidal-risk group and dropout-risk group. In the additional clinical group classification, the results focused more in identifying important features in classifying each group.

4.2.1 Suicidal risk group

1) Feature selection results

Variables	SGL	Lasso	ElasticNet
Q9 Experience of past professional treatment	■	■	■
Q12 Family history of psychosis	■		
Q13 Suicidal thoughts	■	■	■
Housing type	■		
PHQ - Difficulty getting asleep / Sleep too much	■		
PHQ - Tasteless or overeating	■	■	■
PHQ - Being negative to oneself	■		
PHQ - Suicidal/self-injurying thoughts	■	■	■
GAD - Easily get annoyed or angry	■		
PSQI score total	■		
PSQI - Wake up early	■	■	
PSQI - Difficulty in breathing	■	■	■
PSQI - Sleep hours	■		
PSQI - Randomly get up	■	■	
PSQI - Irregular sleep hours		■	
Loss from functional disability	■		
Urge symptoms (Total)		■	
Suicidla impulse symptom	■	■	
Anger symptom	■	■	■
Weight change symptom	■		
Others problems (Total)	■		
Sexual issue problem	■	■	
Mother problem	■		
Death or loss problem	■		
Other gender problem	■		
Personal values problem	■		
Religious problem	■		

Table 5. Common feature selection results in suicidal-risk group classification

As for result of feature selection regarding suicidal risk group, table 5 shows common features selected from experience of past profession help, or features directly related to suicidal thoughts or suffering anger symptom. Selected features include similar features that are reported from prior researches using multi-aspect analysis for predicting suicidal risk. Interestingly, SGL has chosen exceptionally many variables compared to Lasso and ElasticNet. Therefore, how these differences would affect the actual results of the classification was compared afterwards.

2) Classification results using cross validation

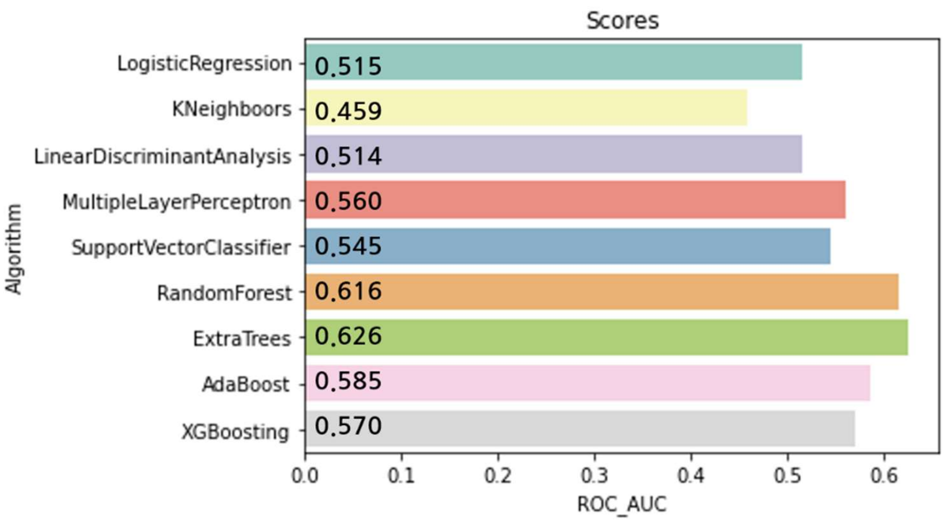


Figure 8. Suicidal-risk group classification results using features selected from Lasso

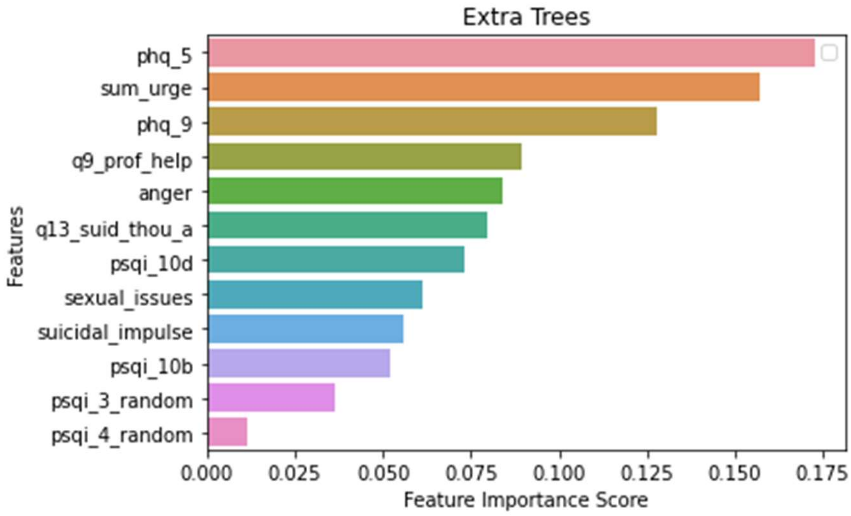


Figure 9. Feature importance from Extra Trees in suicidal-risk group classification

Cross validation results before hyperparameter tuning show low performance compared to the user service group classification. In figure 9, feature importance from ExT were, 'Tasteless or over eating(phq_5)', 'Total sum of urge symptoms(sum_urge)', 'Suicidal/self-injurying thoughts due to depression(phq_9)', 'Experience of past professional treatment(q9_prof_help)', 'Anger symptom(anger)', 'Suicidal thoughts(q13_suid_thou_a)' were selected as importance features in classifying the suicidal risk group.

3) Hyperparameter Tuning

Model Name	CV Score (ROC AUC)	Best Score (ROC AUC)
Logistic Regression	0.515	0.879
K-Nearest Neighbors	0.459	0.882
Linear Discriminant Analysis	0.514	0.865
Support Vector Machine	0.560	0.829
Multi Layer Perceptron	0.545	0.888
Random Forest	0.616	0.861
Extra Trees	0.626	0.873
AdaBoost	0.585	0.867
XGBoost	0.570	0.868

Table 6. Hyperparameter tuning results of suicidal-risk group using features from ElasticNet

As a result of hyperparameter tuning in table 6, model using features selected by ElasticNet resulted best performance. And overall performance of all model significantly increased. However due to radical improvements in models, this may be due to overfitting to target value. Especially, ElasticNet selected minimal number of features which were closely related to suicidal problems. And since the target value of classification was also directly related feature to suicidal risk, and it may have resulted overfitting. So, current results can be used as indicator for assuming users as suicidal group but there is problem with vagueness in target value which will be discussed in following discussion section.

4) Ensemble Modeling

Classifier	**Feature Selecting Method**	**Mean of 100 iterations**
Voting Classifier	Lasso	0.8173
	Sparse-group Lasso	0.8167
	ElasticNet	0.8167
Stacking Classifier	Lasso	0.7793
	Sparse-group Lasso	0.817
	ElasticNet	0.8167

Table 7. Ensemble modeling result of suicidal-risk group classification

Voting	**Predicted Not at risk**	**Predicted Suicidal-risk**
Actual Not at risk	48	1
Actual Suicidal-risk	11	0

Stacking	**Predicted Not at risk**	**Predicted Suicidal-risk**
Actual Not at risk	48	1
Actual Suicidal-risk	11	0

Table 8. Confusion of matrix suicidal-risk group classification using SGL features

Although performance of each model significantly increased after ensemble modeling, the validation through confusion matrix revealed that some models classified all users as not suicidal-risk group which was majority in the total users. Thus, there needed some improvements in ensemble modeling to solve such mis-prediction issues.

4.2.2 Dropout risk group

1) Feature selection results

Variables	SGL	LassoCV	ElasticNet
Q2 Relation satisfaction	■		
PHQ - Feeling no interest	■	■	
PHQ - Tasteless or overeating	■		
PHQ - Being negative to oneself	■		
PSQI - Loud snoring	■	■	
PSQI - Stop breathing	■	■	■
PSQI - Shaking leg	■	■	■
PSQI - Confusion	■		
PSQI - Personal sleep quality	■		
PSQI - Sleep disturbance	■	■	■
Loss from functional disability	■	■	■
Physiological symptoms (Total)	■		
Swelling sensation symptom	■	■	■
Physical symptom	■	■	
Panic symptom	■	■	
Unrealistic feeling symptom	■		
Alcohol abuse symptom	■		
Ataraxia symptom	■		
Insomnia symptom	■		
Club activity problem	■	■	■
Aptitude problem	■		
Anxiety presentation problem	■		
Conflict in family problem	■		
Friend & colleague problem	■		
Ego searching problem	■		
Sibling problem	■		
Living expense problem	■		

Table 9. Common feature selection results in dropout group classification

Table 9 shows that features related to sleep problems(PSQI), daytime functional disabilities, swelling sensation, aptitude problems in club activity were commonly selected as key features in classifying dropout risk group. Also, SGL selected far more features than others and those features were related to aptitude problems, conflicts in relationship. Interesting point is that many prior researches reported counseling service duration, compliance and satisfaction to service as related indicators of dropout, which are features that can be measured during or after the service.(김경원, 박용천, & 김은경, 2020) However, current study selected features from sleep, aptitude and conflict issue for classifying dropout group.

2) Classification results using cross validation

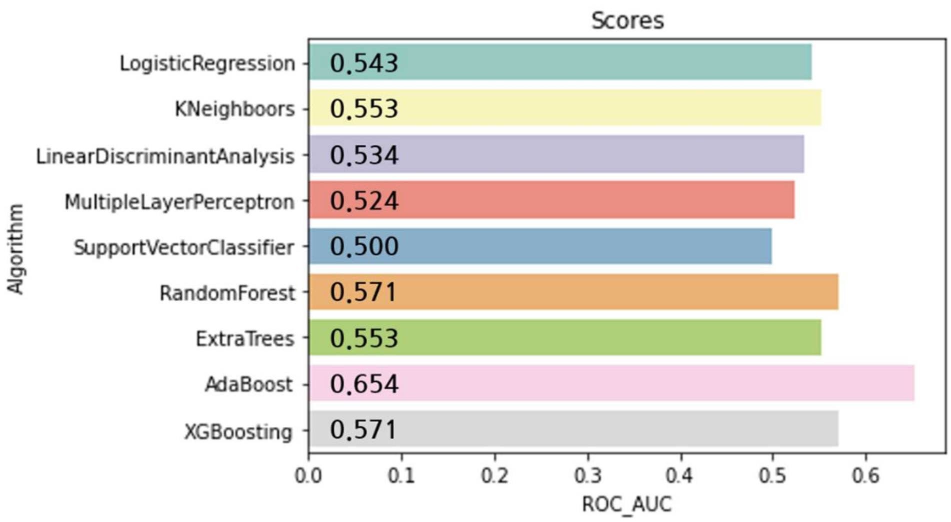


Figure 10. Dropout group classification result using features from SGL

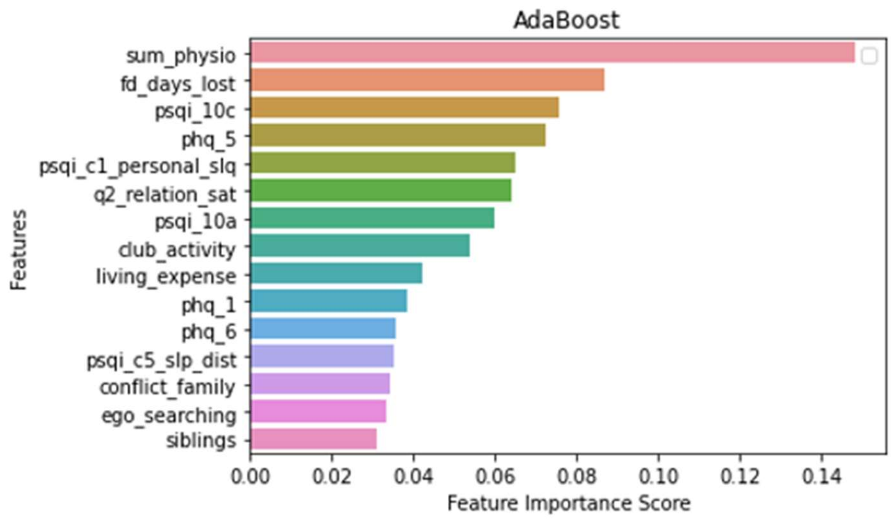


Figure 11. Feature importance from AdaBoost in dropout group classification

As in figure 10, dataset selected by SGL showed best performance. Features related to sleep and relationship problems were considered as important in classifying dropout group. Similar to suicidal-risk group classification, classification in dropout-risk group also showed low performance compared to the user service group classification. Important features selected from AdaBoost in classifying dropout-risk group were 'total sum of physiological symptoms(sum_physio)', 'Day-time function disability(fd_days_lost)', 'Tasteless or overeating due to depression(phq_5)', 'Problems in Personal sleep quality(psqi_c1_personal_slq)', 'Relational satisfaction(q2_relation_sat)', 'More than 30minutes in getting asleep(psqi_10a)', 'Aptitude problem in club activity(club_activity)'.

3) Hyperparameter Tuning

Model Name	CV Score (ROC AUC)	Best Score (ROC AUC)
Logistic Regression	0.543	**0.723**
K-Nearest Neighbors	0.553	0.656
Linear Discriminant Analysis	0.534	0.710
Support Vector Machine	0.524	0.704
Multi Layer Perceptron	0.500	**0.739**
Random Forest	0.571	0.683
Extra Trees	0.553	0.712
AdaBoost	0.654	0.696
XGBoost	0.571	**0.721**

Table 10. Hyperparameter tuning results of dropout group classification using features from Lasso

As a result of hyperparameter tuning in table 10, model using features selected by Lasso resulted best performance. And overall performance of all model significantly increased which was also true in ensemble modeling. By reviewing feature selection methods, selecting adequate number of variables seems to be important in classifying dropout group. This result can be also used as indicator in identifying potential dropout of certain user and help preventing it. However, dropout includes leave of absence and graduation which was not considered in current result. This will be also discussed in future discussion section.

4) Ensemble Modeling

Classifier	Feature Selecting Method	Mean of 100 iterations
Voting Classifier	Lasso	87.25
	Sparse-group Lasso	**88.3**
	ElasticNet	83.5
Stacking Classifier	Lasso	86.48
	Sparse-group Lasso	**88.3**
	ElasticNet	86.05

Table 11. Dropout group ensemble modeling result

Voting	**Predicted Not at risk**	**Predicted Dropout-risk**
Actual Not at risk	52	1
Actual Dropout-risk	7	0

Stacking	**Predicted Not at risk**	**Predicted Dropout-risk**
Actual Not at risk	52	1
Actual Dropout-risk	7	0

Table 12. Confusion of matrix dropout-risk group classification using SGL features

Similar to the suicidal-risk group results, performance of the ensemble modeling greatly increased. However, as in table 12, confusion matrix results indicate that some models classified all dropout group as not-at risk group. Thus, improvements in type 2 error needs to be done.

4.3 Improvements in ensemble modeling and confusion matrix

In the confusion matrix result of user service and clinical group classification, some models classified all users as not-at-risk group if the target group size was small. To solve such issues and improve type 2 errors, adjustments in stacking ensemble classifier was conducted. Selected models were the best performing models from 100 iterations with lowest type 2 error.

4.3.1 User service group

Final Estimator	ROC AUC
K-Nearest Neighbors(KNN)	78.33
Linear Discriminant Analysis(LDA)	75
Support Vector Machine(SVM)	78.33
Multi-Layer Perceptron(MLP)	70
Random Forest(RF)	71.67
Extra Trees(ExT)	76.67
AdaBoost(Ada)	75
XGBoost(XGB)	76.67

Table 13. Refined ensemble modeling results of user service group using SGL features

KNN	**Predicted Counseling**	**Predicted Clinical treatment**
Actual Counseling	29	8
Actual Clinical treatment	5	18

Table 14. Best performing confusion matrix of service group considering type 2 error

By testing various final estimators in stacking classifier, KNN and SVM showed best performance. In addition, type 2 error was lowest in KNN. There were no significant improvements in the user service group classification. Such results may indicate the need of model that can test the effect of each features in classifications rather than adjusting model structure.

4.3.2 Suicidal risk group

There were still many ensemble modeling results classifying all suicidal-risk group as not-at-risk group. In table 15, all models except LDA and XGB predicted suicidal-risk group as not-at-risk group. In addition, as in table 16, type 2 error was still high even in best performing model. This may indicate that better representative feature for the target value is required or other methods to resolve data imbalance may be required.

Final Estimator	ROC AUC
K-Nearest Neighbors(KNN)	81.67
Linear Discriminant Analysis(LDA)	80
Support Vector Machine(SVM)	81.67
Multi-Layer Perceptron(MLP)	81.67
Random Forest(RF)	83.33
Extra Trees(ExT)	81.67
AdaBoost(Ada)	78.33
XGBoost(XGB)	83.3

Table 15. Refined ensemble modeling result of suicidal-risk group using SGL features

LDA	**Predicted Not at risk**	**Predicted Suicidal-risk**
Actual Not at risk	45	4
Actual Suicidal-risk	8	3

Table 16. Best performing confusion matrix of suicidal-risk group considering type 2 error

4.3.3 Dropout group

Final Estimator	ROC AUC
K-Nearest Neighbors(KNN)	90
Linear Discriminant Analysis(LDA)	81.67
Support Vector Machine(SVM)	88.3
Multi-Layer Perceptron(MLP)	88.3
Random Forest(RF)	88.3
Extra Trees(ExT)	88.3
AdaBoost(Ada)	88.3
XGBoost(XGB)	88.3

Table 17. Refined ensemble modeling result of dropout-risk group using ElasticNet features

KNN	**Predicted Not at risk**	**Predicted Dropout-risk**
Actual Not at risk	53	0
Actual Dropout-risk	6	1

Table 18. Best performing confusion matrix of dropout-risk group considering type 2 error

The result of dropout-risk group was similar to the suicidal-risk group. Models except KNN and LDA classified all dropout-risk group as not-at-risk group. Similar approaches considering data imbalance problem and feature representation will help in improving current results.

4.4 User latent group identification

After conducting the user group classification using supervised learning algorithm, the user latent group identification was done using latent class analysis(LCA). First step in the user latent group identification was determining the best number for the model. Then the specific characteristics of each latent group was analyzed using validated features.

4.4.1 Latent Class Analysis

Class number	**Iteration count**	**BIC**	**Model distribution**						
			G1	G2	G3	G4	G5	G6	G7
2	105	-40404.23	51 (0.17)	249 (0.83)	-	-	-	-	-
3	163	-43183.30	249 (0.83)	29 (0.0.097)	22 (0.073)	-	-	-	-
4	7	-48527.38	115 (0.383)	126 (0.42)	26 (0.087)	33 (0.11)	-	-	-
5	**35**	**-53817.21**	**25 (0.083)**	**130 (0.433)**	**22 (0.0733)**	**119 (0.397)**	**4 (0.013)**	**-**	**-**
6	68	-47453.09	25 (0.083)	26 (0.0867)	37 (0.114)	104 (0.356)	0 (0.0)	108 (0.36)	-
7	129	-53534.01	4 (0.013)	11 (0.037)	119 (0.397)	11 (0.037)	130 (0.433)	9 (0.03)	16 (0.053)

Table 19. Changes in latent class analysis results according to the number of classes

Table19 shows the difference of model according to the number of classes. As a result of comparison using BIC, the model with five classes was chosen to be the most suitable. Since BIC is an indicator in comparing between models, and it cannot represent the actual performance of the model itself.

Therefore, validation procedure for each categorical and numerical variables was done to test how each features effected in distinguishing each latent groups.

4.4.2 Features validation after latent class analysis

Variable	P-value	Cramer's V
PHQ - Category	1.99E-05***	0.219389
GAD - Category	0.002***	0.184123
PSQI - Category	0.001***	0.242404
Q9 Experience of past professional treatment	0.012***	0.206908
Difficult focusing symptom	0.007***	0.217536
Decreased memory symptom	0.001***	0.242249
Difficult in dailylife symptom	0.006***	0.188309
Alcohol abuse symptom	0.007***	0.216524
Insomnia symptom	6.62E-05***	0.285219
General health problem	0.018***	0.199695
Unrealistic feeling symptom	0.008***	0.214624
Senior/junior relationship problem	0.013***	0.205162
Sexual issues problem	0.014***	0.20411
Service category	4.42E-05***	0.290272

Table 20. Validation result of categorical features using Cramer's V

Variable	F-value	Pr(>F)	partial η^2
PHQ - Tasteless or overeating	5.205	0.0004***	0.0659
PHQ - Being negative to oneself	5.645	0.0002***	0.0711
PHQ - Difficulty in focusing	7.146	1.67E-05***	0.0883
PHQ - Difficuly due to depressive symptoms	6.924	2.44E-05***	0.0858
PHQ - Total score	8.577	1.47E-06***	0.1041
GAD - Overwhelming fear	6.324	6.79E-05***	0.0790
PSQI - Personal sleep quality	6.372	6.26E-05***	0.0795
PSQI - Taking sleep-related drugs	318.4	<2e-16***	0.8119
PSQI - Other sleep problems	487	<2e-16***	0.8685
PSQI - Funtional disability due to sleep issue	43.94	<2e-16***	0.3733
PSQI - Stop breathing	5.394	0.0003***	0.0682
PSQI - Total score	18.74	9.71E-14***	0.2026
Funtional disability total score	5.74	0.0001***	0.0722
Days loss due to functional disability	5.737	0.0002***	0.0722

Table 21. Validation result of numeric features using partial η^2

Table 20 and 21 shows the validation results of categorical and numeric values. Both tables only contain features that were determined as valid and table 21, which is result of numeric features, only contains feature that are both significant and partial η^2 over 0.06 meaning over moderate effect in classifying each latent groups. For categorical values, categorical features about the user service group, depression(PHQ), anxiety(GAD), sleep issues(PSQI), symptoms and problems from various areas. Also, past experience of professional help was selected as valid in identifying each latent group. For numeric features, various features from severity scores related to depression and anxiety, sleep problems and functional disability were chosen as valid. Features from various areas were included similar to the user service group classification results. By using these validated features, interpretation of individual latent group characteristic was done.

4.4.3 Analysis of individual user latent group characteristics

After validating features for identifying each latent group, comparison between each group using such features were conducted. The first group tested was group 3, which showed higher problematic levels compared to other group and labeled as high-risk group. Next, group 5 showed highest sleep problems and labeled as sleep problem group. Group 5 was thought to be sub-group of group 3, but showed relatively lower severity and divided into different group. Group 1 and 2 showed similar results in survey scores, while ratio of treatment group and high-risk group in depression(PHQ) and anxiety(GAD) was higher in group 1. So, group 2 was label as ‘lower-risk group’ and group 1 as ‘lower/moderate group’. Group 2 and 4 show difficulty in similar areas while group 4 contains for higher ratio of treatment group, high-risk group in depression and anxiety, suffering difficulties, and higher severity in problem(sexual issue). Also Group 4 was suspected as potential dropout group due to group’s suffering problem and symptoms.

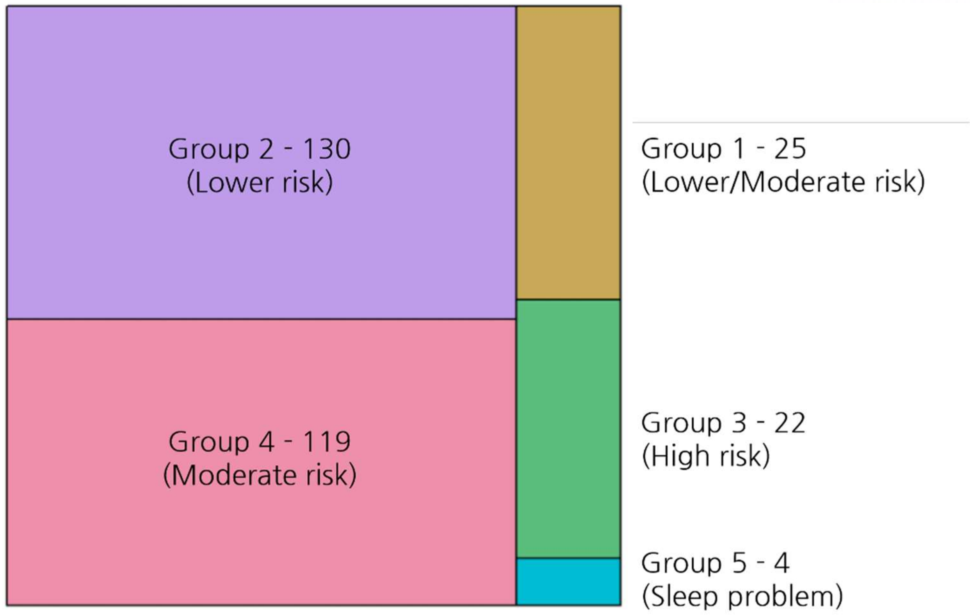


Figure 12. Distribution of all groups among total users

Figure 12 shows the treemap plot of distribution of all groups. Lower risk group holds largest proportion of users and followed by moderated risk group. Lower/moderate risk group and high-risk group holds similar number of users and sleep problem group holds only 4 users in total. Individual group characteristics are described as below. Specific analysis of each latent group is as followed.

1) Lower risk group(Group2)

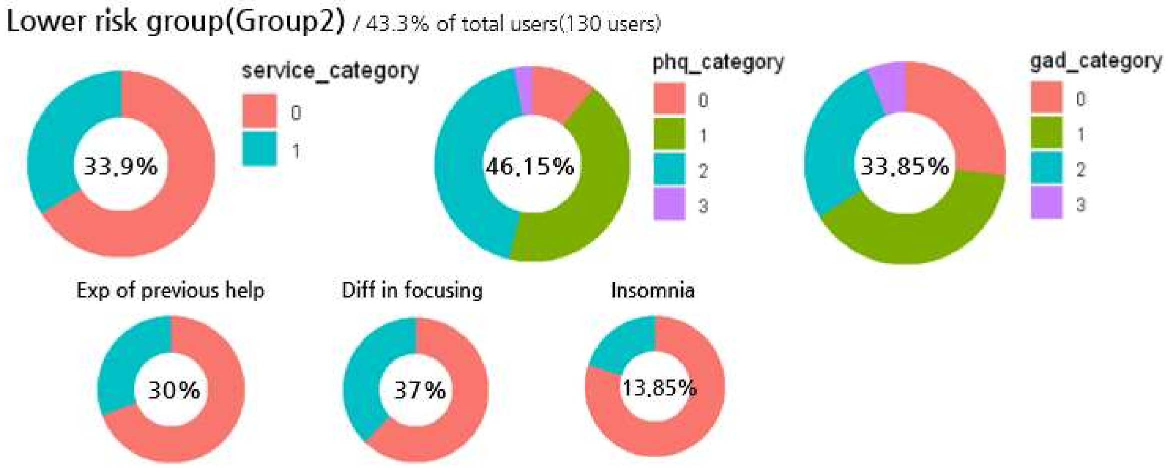


Figure 13. Individual characteristics of lower risk group

First group is lower-risk group. About 34% of lower-risk group belonged to clinical treatment group. Ratio of higher-risk group meaning moderate/high severity in depression and anxiety was 46% and about 34% each. This group reported lowest severity results apart from group 5(sleep problem group). Users experienced professional help before was 30% and suffered difficulty in focusing(37%).

2) Moderate-risk group(Group4)

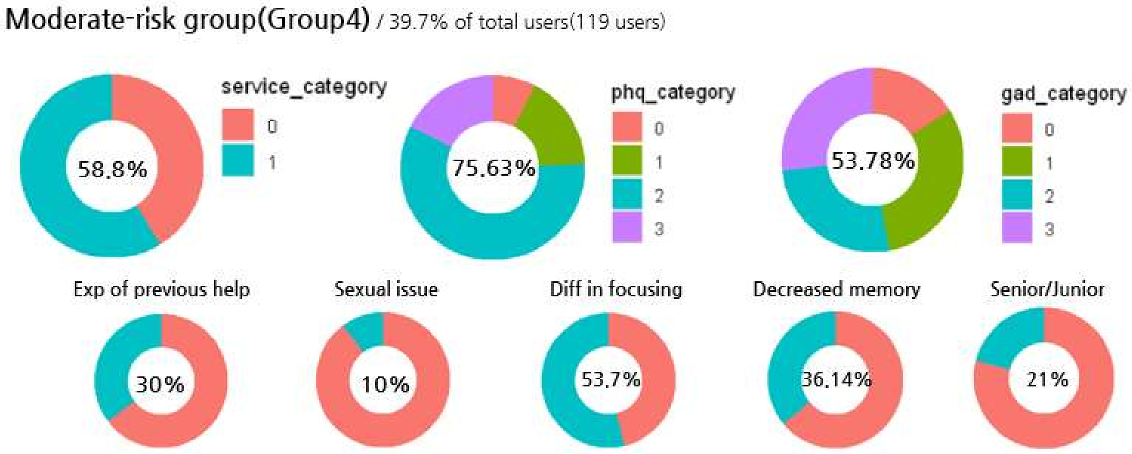


Figure 14. Individual characteristics of moderate risk group

Second is moderate-risk group. About 59% of users were in clinical treatment group and relatively higher proportion of depression and anxiety higher-risk group was present with about 75% and 53% each. Difficulty due to sexual issue was highest among all groups while 35% experienced professional help before. Difficulty in focusing(53.7%) was similar to high-risk group and decreased memory(36.14%) was higher than high-risk group. Various symptoms such as difficulty in daily life(17.7%), alcohol abuse(6.72%), unrealistic feelings(14.29%) were also reported higher compared to other groups. Also problem in senior/junior relationship(21%) was reported to be highest among all groups.

3) Lower/moderate risk group(Group1)

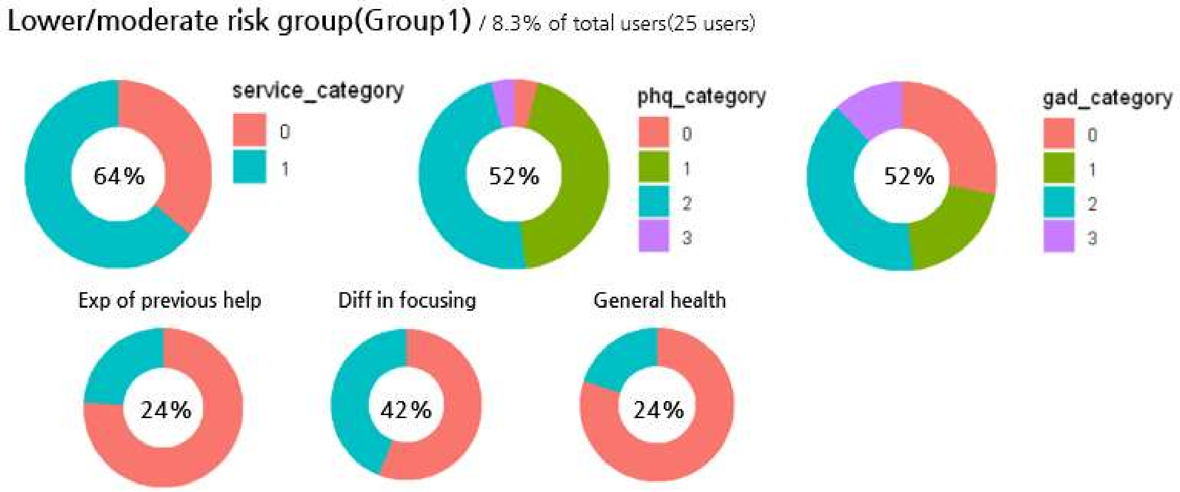


Figure 15. Individual characteristics of lower/moderate risk group

Third group is lower/moderate risk group. Ratio of clinical treatment group was 64% among group lower/moderate risk group. Ratio belonging to higher-risk group from depression(PHQ) and anxiety(GAD) category was 52%. Users experienced professional help before was 24% and suffered difficulty in focusing(42%). 20% of users reported insomnia and 24% reported general health problems.

4) High-risk group with sleep issue(Group3)

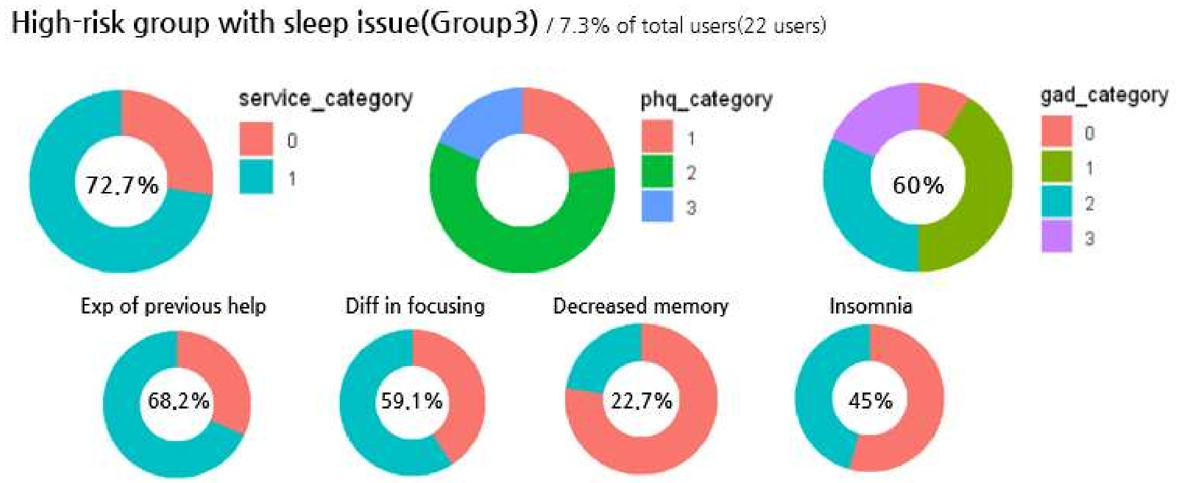


Figure 16. Individual characteristics of higher risk group with sleep issues

Next is high-risk group with sleep issue. All users in group 3 reported sleep problems. 72.7% were clinical treatment group with 77.27% and 60% in depression and anxiety high-risk group. 68.2% of users reported experience of past professional help rating highest ratio among all other groups. Cognitive and behavior sufferings were difficulty in focusing, decreased memory, difficulty in daily life showing higher proportion compared to many other groups. Users suffering insomnia was 45% with report of taking sleep-related drugs which was found only in high-risk with sleep issue and sleep-problem group.

5) Sleep problem group(Group5)

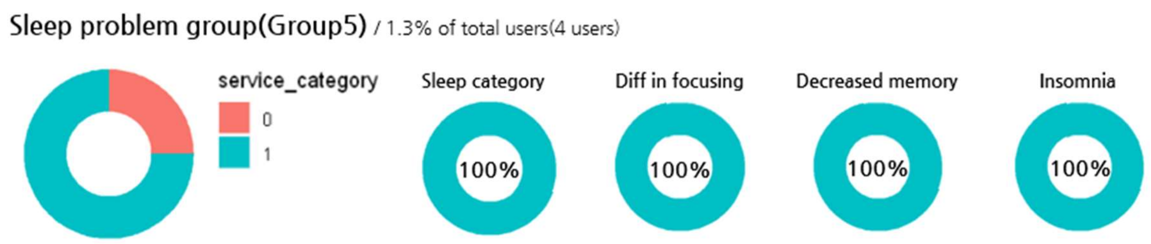


Figure 17. Individual characteristics of sleep problem group

Last group is sleep problem group. This group was smallest, but most distinct group. Sleep problem group showed highest sleep(PSQI) scores with high score in suffering personal sleep quality, daily functional disability due to sleep problem, suffering other sleep problems, taking sleep-related drugs. All users reported difficulty in focusing, decreased memory, insomnia.

V. DISCUSSION

The purpose of current study was to analyze information about users which is hard to detect at the initial stage of service based on service application data and to help clinicians to improve service effectiveness. First, the user service group classification, which classify two representative services – counseling and clinical treatment, was conducted. Through this analysis, various models were applied and important variables and models for classifying service groups were confirmed. In addition, by suggestion from counselor feedback, classification of two additional clinical groups, suicidal-risk and dropout-risk group, was conducted. Latent class analysis was used to search for potential user groups that are difficult to identify with original clinical labels. As a result, five sub-groups were identified and provided detailed group statistical information, allowing clinicians to effectively identify characteristics of users visiting university counseling centers. The specific contribution of current study in improving university counseling center service is as follows.

First, current results can be applied in analyzing service effectiveness and providing customized services. Current study successfully identified various features for classifying user service group. Depression, suicidal-risk, sleep problems were related to service group classification. Comparing service effectiveness depending on classified service group and actual service provided can be done. Providing customized service based on to such analysis will enhance overall service effectiveness provided to users. Moreover, service effectiveness regarding predicted service group can be done. Such analysis will measure the effectiveness of proper clinical classification and intervention and how counseling centers can improve their service. Still, identifying features with ML algorithm alone was not good enough. Collaborating with qualitative method will increase perfection of the results.

In addition, current result can be used as reference for converting service into clinical treatment and early intervention. In some cases, it is very difficult to determine between two service categories. For example, there are some cases of high-risk group assigned as counseling due to refusal of drug use in service. In such cases, current results can be used as reference for future converting to clinical treatment. For example, counseling centers without co-operated psychiatric clinics can determine external clinical treatment for their users. This will allow users in need of drug-usage to receive appropriate service at early stage to get higher service effectiveness.

As for classifying suicidal risk group, current study successfully identified similar features from prior researches using ML algorithm. Suicidal thoughts, physiological symptoms, anger symptom was selected as key features in classifying suicidal risk group. Prior research using multiphasic inventory(MMPI-2-R) reported similar features as effective in classifying suicidal risk and current study made its achievement in identifying similar features by using initial stage service data so that counselors in university counseling centers can easily receive similar results without additional work or burden.

Still, there are some considerations in feature selection methods and model result. There was difference between selected features and possible overfitting was detected. Better clinical label for valid target value may be required. Qualitative feature selection collaborating with counselors may be effective in classifying and managing target group.

Dropout-risk group classification identified features that were not found in prior research. Prior studies using qualitative methods reported that detecting dropout group is difficult and the criteria is also vague. Prior studies suggested counseling durations, satisfaction and compliance as dropout indicators which were hard to identify at initial stage the service. Physiological symptoms, sleep problems, relationship problems were selected as importance features in classifying dropout group. In addition, dropout group suffer both aptitude and psychological problems, possibly indicating sleep and functional disabilities due to aptitude issues. As a result, taking such features into account in providing service will help in reducing dropout and providing continuous service to users.

Dropout group also had difference between feature selection method. However, commonly selected features were sleep and adaptive problems and other non-common features were part of aptitude problems. Thus, further study comparing two methods first focusing on few key features and exploring various feature will improve overall understanding on dropout issues. In addition, dropout include early terminations due to leave of absence and graduation. Applying such background information to the service records will provide better representation of user groups.

Current study conducted explorative research using various feature selection methods and machine learning models. As a result, current results identified features and models that successfully identified the user service and clinical group. Service group classification had higher similarity between feature selection methods. Among the feature selection methods, SGL showed relatively higher ROC AUC with lower type 2 error. Still, there were difference between each selection method and model that can test the difference according to changes in selected features is required. For machine learning models, boosting models showed better performance compared to relatively simpler models. Thus, model optimization based on current data is expected to enhance the classification performance.

In suicidal-risk group, feature selection methods which selected more features like SGL and Lasso showed better performance. In addition, simpler models like KNN and LDA showed better results compared to other models meaning there might be distinctive difference between suicidal-risk and normal group. However, there is possible improvement in that many models did not identified suicidal-risk group properly. Therefore, it is necessary to find and manage features that can better reflect the characteristics of suicidal-risk group.

Feature selection method in dropout-risk group which selected particularly many features barely identified dropout-risk correctly. Still, such tendency was also present in Lasso and ElasticNet which selected small number of features. This may be due to the small number of target group and variety within group characteristics. Additional efforts using anomaly detection or resampling methods to solve data imbalance issues will be required as following.

By conducting latent user group identification for users visiting university counseling center, current study identified sub-groups that can be applicable to existing counseling service. Categorical variables showed differences between latent groups in various areas such as depression, anxiety, sleep issue, and suffering problems. Also, the service group showed differences, and it was confirmed that there was a difference in severity between each potential user group. Especially, by using such results, it is possible to detect the deterioration of the symptoms and developing to higher risk group in advance. Numerical variables showed differences in various aspects such as depression, anxiety, sleep issue, and functional disability. Especially, many variables were selected from depression and sleep severity scores. Like categorical variables, the numerical severity each group is suffering can used as an indicator to objectively identify the service effectiveness of users.

Total 5 latent group was identified using validated variables and showed similar classifications to clinical practice and identified additional special group. As a result of analysis of each subgroup, low/moderate-risk group, lower-risk group, high-risk groups with sleep problems, moderate-risk groups, and sleep problem group were identified. As with the previous user service group classification, the service effectiveness analysis according to the corresponding classification based on the service record will contribute in providing adequate service to each user. Such results can be applicable as separate criteria for individual user's severity apart from user service and clinical group which allows more accurate understanding to users.

In addition, by using data from different university counseling centers, exploring difference in latent group characteristics according to the university will be possible. Such approach will help individual counseling centers to provide customized service according to user characteristics and prepare for possible new latent groups. Especially, it is thought that current result can be applied in a way of focusing on specific demand such as grades, career, and adaptation which can expand the area of service provided to users.

Also, suggesting possible clinical approaches according to individual latent groups is possible. Approaches like starting early intervention to users detected as high-risk group to prevent from getting worse and constant tracking and specification of reporting problems in moderate-risk group to prevent dropout and developing to high-risk group will help maintain user symptom severity more effectively.

In addition, focusing on sleep-related needs distinguished from high-risk group for effective service can result more effective problem solving to sleep issue group. By applying such approaches according to each latent group, service effectiveness will be improved.

5.1 Further research

This study successfully proposed suggestions in reducing the burden of counselors at the early stage of service. In addition, effective information about future service planning was derived using current study results. Several further researches are expected to expand the effectiveness of current study.

First is development of refined data management system that can reflect user characteristics better and improve the performance of machine learning models using such methods. The data used in this study is not collected for research purpose but for service management. Although questionnaires are included that are used in actual screening, there are more potentials for some improvement. In particular, for dropout-risk group, leave of absence is an important factor, but it is difficult to handle such information with current system. Therefore, analyzing clinical groups that needs continuous monitoring and developing system for detecting group characteristic will result effective collection of data from the beginning. Especially, if a model that can test the effect of certain feature to the user group classification is developed, validating efficiency of data collection system and building more accurate machine learning model will be possible.

Second, it is possible to analyze the service effectiveness according to the user group classification by utilizing the entire service record based on the improved data collection system and model. Service effectiveness analysis has the advantage that it is possible to confirm not only effectiveness of the service but also the side effects caused by the wrong service provision. The service plan considering user group classification will help to increase the service effect of the university counseling center.

Finally, applying current study researches to other university counseling centers can be done. Current study analyzed the users visiting UNIST healthcare center. Thus, testing refined machine learning model and classification methods to other university counseling centers will derive unique characteristics of each school. Such attempts will enable counseling centers to detect their visitor characteristics easily and provide customized service according to different user needs. The application to actual services is expected to help the university counseling centers to manage their users more effectively, especially in the case of a small counseling centers without affiliated psychiatric clinics.

REFERENCE

Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry research, 28*(2), 193-213.

Cho, G., Yim, J., Choi, Y., Ko, J., & Lee, S.-H. (2019). Review of machine learning algorithms for diagnosing mental illness. *Psychiatry investigation, 16*(4), 262.

Flynn, C., & Heitzmann, D. (2008). Tragedy at Virginia Tech: Trauma and its aftermath. *The Counseling Psychologist, 36*(3), 479-489.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine, 16*(9), 606-613.

Kwan, M., Arbour-Nicitopoulos, K., Duku, E., & Faulkner, G. (2016). Patterns of multiple health risk–behaviours in university students and their association with mental health: application of latent class analysis. *Health promotion and chronic disease prevention in Canada: research, policy and practice, 36*(8), 163.

Leon, A. C., Olfson, M., Portera, L., Farber, L., & Sheehan, D. V. (1997). Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *The international journal of psychiatry in medicine, 27*(2), 93-105.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research, 12*, 2825-2830.

Seo, J.-G., Cho, Y. W., Lee, S.-J., Lee, J.-J., Kim, J.-E., Moon, H.-J., & Park, S.-P. (2014). Validation of the generalized anxiety disorder-7 in people with epilepsy: a MEPSY study. *Epilepsy & Behavior, 35*, 59-63.

Sheehan, D. (1983). : The Anxiety Disease. New York, E. *NY: Charles Scribner Sons*.

Simon, N., Friedman, J., Hastie, T., Tibshirani, R., & Simon, M. N. (2018). Package 'SGL'. *CRAN Documentation*.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine, 166*(10), 1092-1097.

Tate, A. E., McCabe, R. C., Larsson, H., Lundström, S., Lichtenstein, P., & Kuja-Halkola, R. (2020). Predicting mental health problems in adolescence using machine learning techniques. *PloS one, 15*(4), e0230389.

Visser, I., & Speekenbrink, M. (2016). Dependent mixture models—hidden Markov models of GLMs and other distributions in S4. *R package version 1.3-3*.

건강보험심사평가원. (2018, 2018-12-13). 연령대별 많이 나타나는 정신건강 질환은? Retrieved from http://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA020041000100&brdScnBltNo=4&brdBltNo=9731

김경원, 박용천, & 김은경. (2020). 정신과 외래 진료 환자들의 MMPI-2-RF 타당도 척도와 치료 조기 종결 및 치료 지속 기간 간의 관계. *생물치료정신의학, 26*(1), 56-64.

김은하, 전소연, & 김다예. (2016). 대학상담센터의 자살예방과 개입에 대한 현황 및 실태 조사. *인간이해, 37*(1), 1-20.

노윤신, & 이성은. (2019). *대학생 정신건강 실태와 심리상담 지원의 쟁점 및 과제*. Retrieved from

박준영, & 김지혜. (2010). 한국판 Sheehan Disability Scale 의 신뢰도와 타당도 연구. *Korean Journal of Clinical Psychology, 29*(1), 73-81.

안제용, 서은란, 임경희, 신재현, & 김정범. (2013). 한국어판 우울증 선별도구 (Patient Health Questionnaire-9, PHQ-9) 의 표준화 연구. *생물치료정신의학, 19*(1), 47-56.

이영은, 차영은, & 민경화. (2013). 대학상담소 접수면접 체제에 대한 개선방안 연구. *인간이해, 34*(2), 1-19.

이혜선, 김성연, 박일, 강여정, 이지영, & 권정혜. (2012). 대학생의 자살관련생각과 행동의 원인 및 자살을 선택하지 않은 이유. *한국심리학회지: 상담 및 심리치료, 24*(3), 703-728.

ACKNOWLEDGEMENT

During my master's degree, many people have supported me and gave many insights to go one step further. Without them, I would not have completed this research and master's degree. I would like to thank the following people with all my heart.

First, I would like to express my deepest gratitude to my advisor Prof. Dooyoung Jung. I was able to complete my degree under his constant trust and support. Choosing Prof. Dooyoung Jung as my advisor was one of the best things that I did after entering UNIST. Also, I would like to thank Prof. Sung Phil Kim and Prof. Chiehyeon Lim for taking the role as my thesis committee. Their warm attention and comments enable me to successfully complete my thesis. In addition, I would like to thank all the members of Healthcare Analytics and Interface lab. Dr. Sangil Lee always gave me insightful advice so that I can successfully conduct my research. Dokyung Kim, Sunmi Lee and Jiwoo Lim always gave me insights and motivations throughout my degree.

Secondly, I would like to thank my fellow friends, Joontae Ki, Changjin Kim, Sangsuk Lee, Sungwon Lee, Chanwoo Jung, Younghwan Jun, who always stood by my side and gave me strength to carry on.

Finally, I would like to thank my parents and sister for their unconditional support and love letting me to proceed with my degree. It would not have been possible without the dedication and sacrifice that they have given to me.

APPENDIX

Figure A1. Feature importance result in service group classification using total pre-processed data

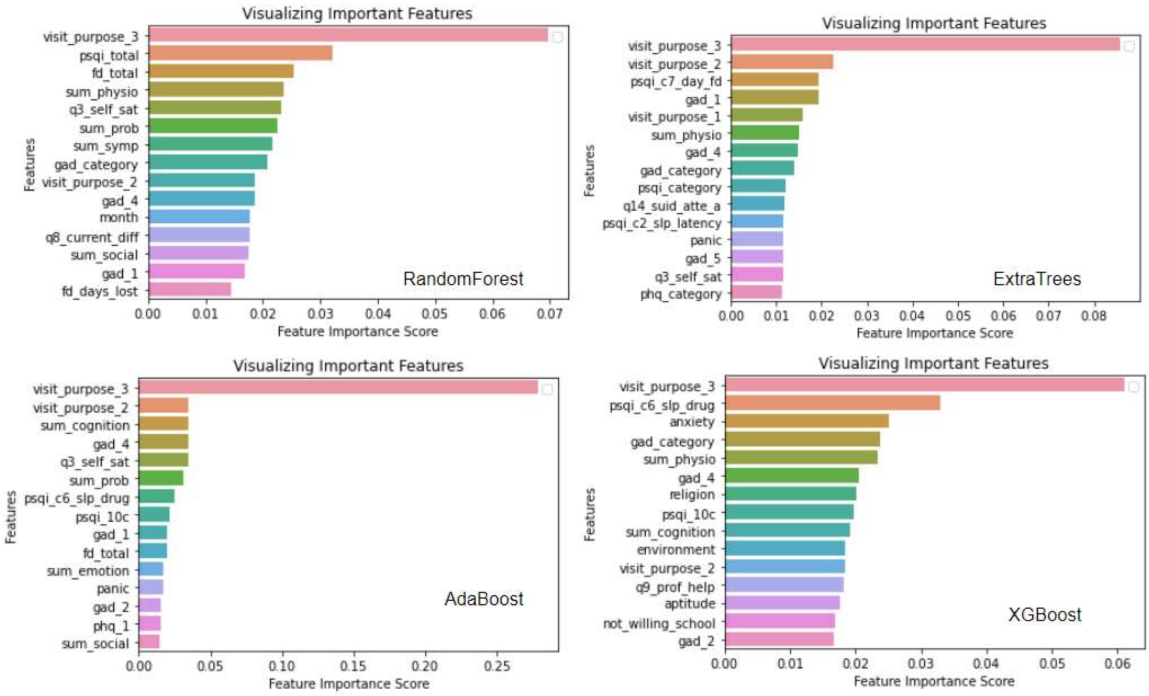
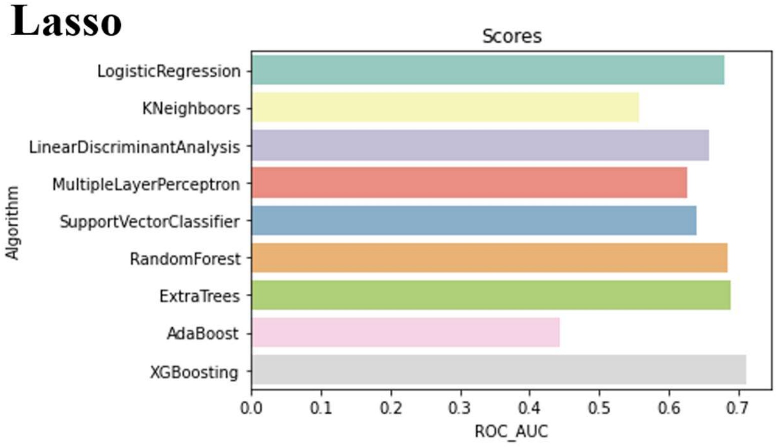


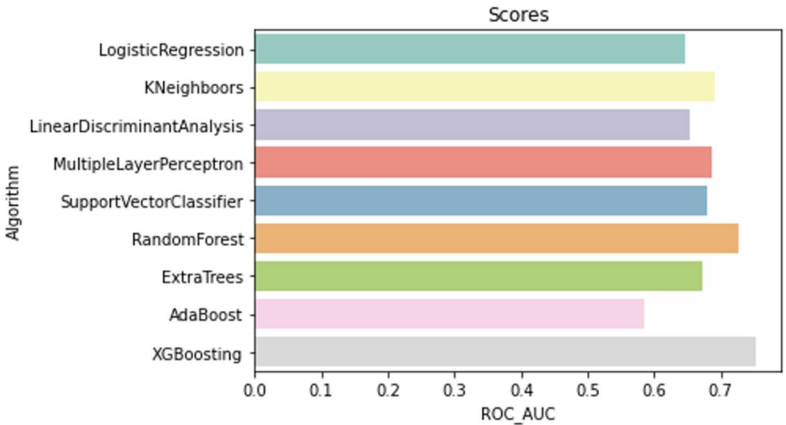
Figure A2. Cross-validation result in service group classification using Lasso/ElasticNet features

Lasso



	ROC_AUC	Algorithm
0	0.680376	LogisticRegression
1	0.558167	KNeighboors
2	0.658637	LinearDiscriminantAnalysis
3	0.626322	MultipleLayerPerceptron
4	0.639835	SupportVectorClassifier
5	0.685664	RandomForest
6	0.688602	ExtraTrees
7	0.444183	AdaBoost
8	0.712691	XGBoosting

ElasticNet



	ROC_AUC	Algorithm
0	0.645123	LogisticRegression
1	0.690952	KNeighboors
2	0.653349	LinearDiscriminantAnalysis
3	0.685664	MultipleLayerPerceptron
4	0.680376	SupportVectorClassifier
5	0.726204	RandomForest
6	0.672150	ExtraTrees
7	0.585194	AdaBoost
8	0.753231	XGBoosting

Figure A3. Feature importance result in service group classification using SGL features

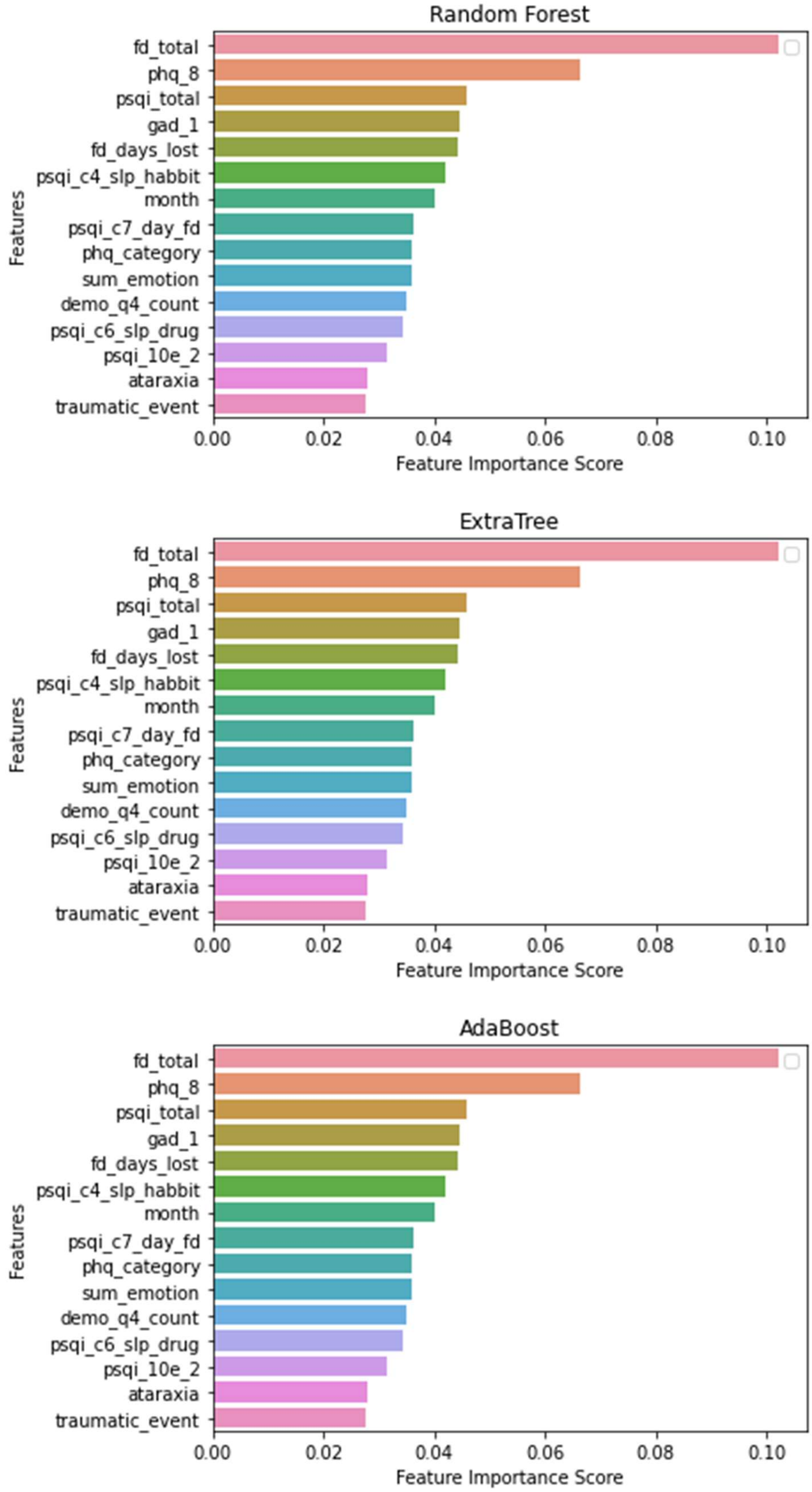


Table A1. Hyperparameters used for each models

Model Name	**Hyperparameters**
Logistic Regression	'penalty': 'l1', 'l2', 'elasticnet', 'none' 'C': logspace(-4,4,20) 'solver': 'lbfgs', 'newton-cg', 'liblinear', 'sag', 'saga' 'max_iter': 100, 1000, 2500, 5000
K-Nearest Neighbors	'n_neighbors': 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 'weights' : 'uniform', 'distrance' 'metirc': 'eucliden', 'manhattan', 'mahalanobis' 'algorithm': 'auto', 'ball_tree', 'kd_tree', 'brute'
Linear Discriminant Analysis	'solver': 'svd', 'lsqr', 'eigen' 'shrinkage': 'auto', 'float', 'none'
Support Vector Machine	'kernel': 'rbf' 'gamma' : 0.001, 0.01, 0.1, 1 'C': 1, 10, 50, 100,200,300, 1000
Multi Layer Perceptron	'hidden_layer_sizes': (50,50,50), (50,100,50), (100,) 'activation': 'tanh', 'relu' 'solver': 'sgd', 'adam' 'alpha': 0.0001, 0.05, 'learning_rate': 'constant', 'adaptive'
Random Forest	'max_depth': None 'max_features': 1, 3, 10 'min_samples_split': 2, 3, 10 'min_samples_leaf': 1, 3, 10 'bootstrap': False 'n_estimators' : 100,300 'criterion': 'gini'
Extra Trees	'max_depth': None 'max_features': 1, 3, 10 'min_samples_split': 2, 3, 10 'min_samples_leaf': 1, 3, 10 'bootstrap': False 'n_estimators' : 100,300 'criterion': 'gini'
XGBoost	'min_child_weight': 1, 5, 10 'gamma': 0.5, 1, 1.5, 2, 5 'subsample': 0.6, 0.8, 1.0 'colsample_bytree': 0.6, 0.8, 1.0 'max_depth': 3, 4, 5

Table A2. Hyperparameters selected at user group classification

LassoCV

Model Name	Best Score (ROC AUC)	Best Hyperparameters
Logistic Regression	0.756	{'C': 0.0001, 'max_iter': 100, 'penalty': 'l2', 'solver': 'lbfgs'}
K-Nearest Neighbors	0.724	{'algorithm': 'auto', 'metric': 'manhattan', 'n_neighbors': 15, 'weights': 'distance'}
Linear Discriminant Analysis	0.742	{'shrinkage': 'auto', 'solver': 'lsqr'}
Support Vector Machine	0.767	{'C': 1, 'gamma': 0.1, 'kernel': 'rbf'}
Multi Layer Perceptron	0.773	{'activation': 'relu', 'alpha': 0.0001, 'hidden_layer_sizes': (50, 50, 50), 'learning_rate': 'adaptive', 'solver': 'sgd'}
Random Forest	0.754	{'bootstrap': False, 'criterion': 'gini', 'max_depth': None, 'max_features': 1, 'min_samples_leaf': 3, 'min_samples_split': 10, 'n_estimators': 300}
Extra Trees	0.776	{'bootstrap': False, 'criterion': 'gini', 'max_depth': None, 'max_features': 1, 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 100}
XGBoost	0.711	{'colsample_bytree': 0.6, 'gamma': 1, 'max_depth': 3, 'min_child_weight': 1, 'subsample': 0.8}

SGL

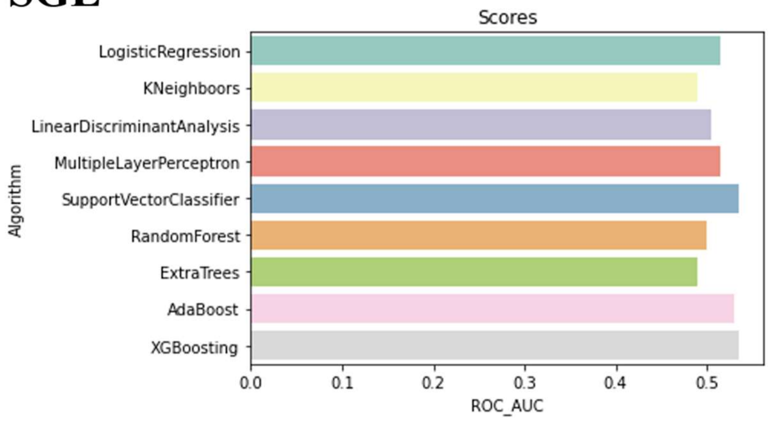
Model Name	Best Score (ROC AUC)	Best Hyperparameters
Logistic Regression	0.751	{'C': 0.0001, 'max_iter': 100, 'penalty': 'l2', 'solver': 'lbfgs'}
K-Nearest Neighbors	0.738	{'algorithm': 'brute', 'metric': 'mahalanobis', 'n_neighbors': 17, 'weights': 'distance'}
Linear Discriminant Analysis	0.730	{'shrinkage': 'auto', 'solver': 'lsqr'}
Support Vector Machine	0.766	{'C': 1, 'gamma': 0.01, 'kernel': 'rbf'}
Multi Layer Perceptron	0.774	{'activation': 'relu', 'alpha': 0.05, 'hidden_layer_sizes': (50, 100, 50), 'learning_rate': 'adaptive', 'solver': 'sgd'}
Random Forest	0.750	{'bootstrap': False, 'criterion': 'gini', 'max_depth': None, 'max_features': 1, 'min_samples_leaf': 10, 'min_samples_split': 2, 'n_estimators': 100}
Extra Trees	0.768	{'bootstrap': False, 'criterion': 'gini', 'max_depth': None, 'max_features': 1, 'min_samples_leaf': 3, 'min_samples_split': 10, 'n_estimators': 100}
XGBoost	0.721	{'colsample_bytree': 1.0, 'gamma': 2, 'max_depth': 5, 'min_child_weight': 1, 'subsample': 1.0}

ElasticNet

Model Name	Best Score (ROC AUC)	Best Hyperparameters
Logistic Regression	0.751	'C': 0.0006951927961775605, 'max_iter': 100, 'penalty': 'l2', 'solver': 'liblinear'
K-Nearest Neighbors	0.722	'algorithm': 'auto', 'metric': 'manhattan', 'n_neighbors': 19, 'weights': 'distance'
Linear Discriminant Analysis	0.738	'shrinkage': 'auto', 'solver': 'lsqr'
Support Vector Machine	0.754	'C': 1, 'gamma': 0.01, 'kernel': 'rbf'
Multi Layer Perceptron	0.764	'activation': 'relu', 'alpha': 0.05, 'hidden_layer_sizes': (50, 50, 50), 'learning_rate': 'constant', 'solver': 'sgd'
Random Forest	0.739	'bootstrap': False, 'criterion': 'gini', 'max_depth': None, 'max_features': 1, 'min_samples_leaf': 3, 'min_samples_split': 10, 'n_estimators': 100
Extra Trees	0.765	'bootstrap': False, 'criterion': 'gini', 'max_depth': None, 'max_features': 1, 'min_samples_leaf': 3, 'min_samples_split': 10, 'n_estimators': 300
XGBoost	0.712	'colsample_bytree': 1.0, 'gamma': 1, 'max_depth': 4, 'min_child_weight': 1, 'subsample': 1.0

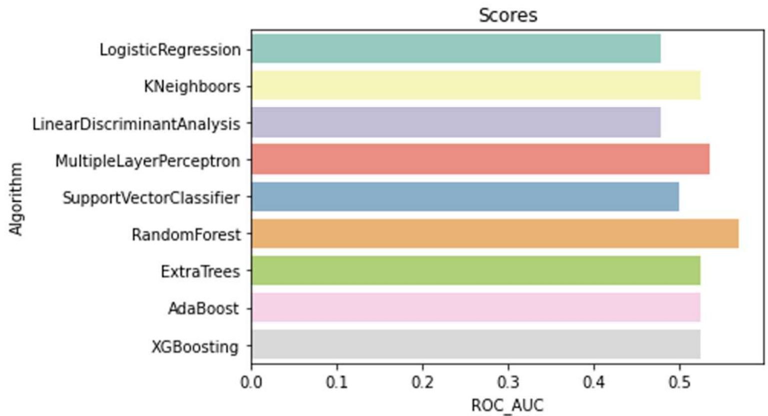
Figure A4. Cross-validation result in suicidal-risk group classification using SGL/ElasticNet features

SGL



	ROC_AUC	Algorithm
0	0.514842	LogisticRegression
1	0.489796	KNeighboors
2	0.504638	LinearDiscriminantAnalysis
3	0.514842	MultipleLayerPerceptron
4	0.535250	SupportVectorClassifier
5	0.500000	RandomForest
6	0.489796	ExtraTrees
7	0.529685	AdaBoost
8	0.535250	XGBoosting

ElasticNet



	ROC_AUC	Algorithm
0	0.479592	LogisticRegression
1	0.525046	KNeighboors
2	0.479592	LinearDiscriminantAnalysis
3	0.535250	MultipleLayerPerceptron
4	0.500000	SupportVectorClassifier
5	0.570501	RandomForest
6	0.525046	ExtraTrees
7	0.525046	AdaBoost
8	0.525046	XGBoosting

Figure A5. Feature importance result in service group classification using Lasso features

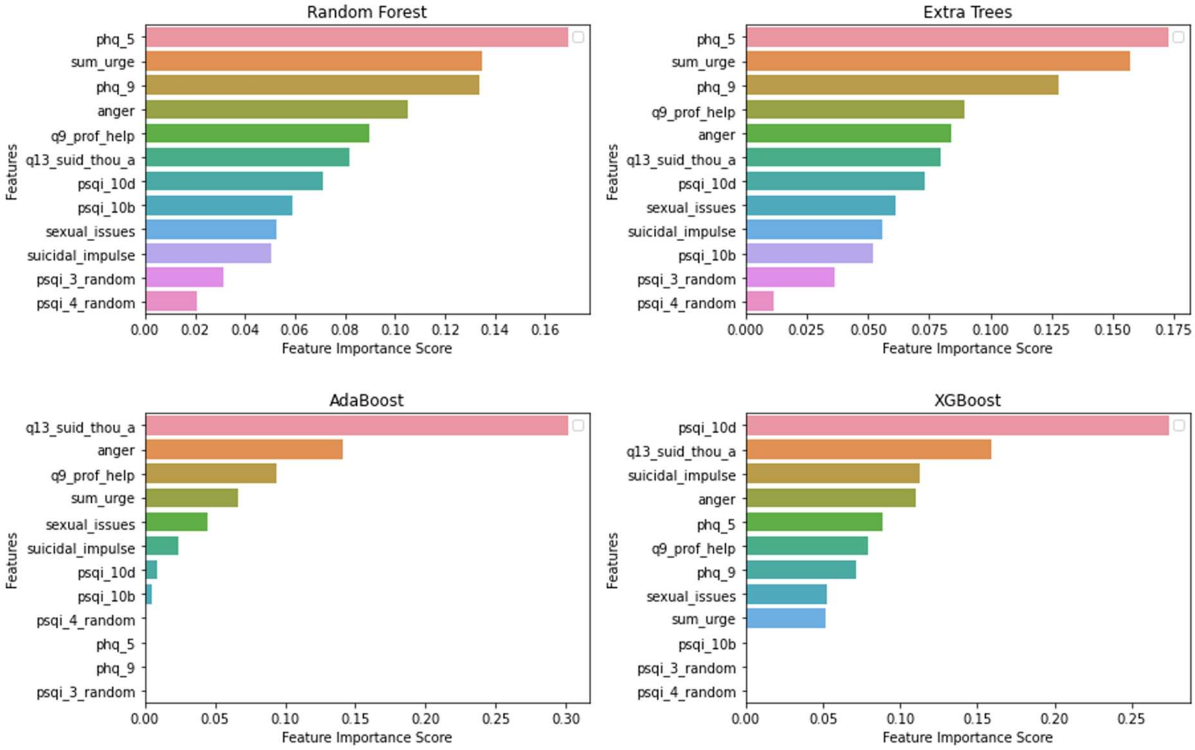
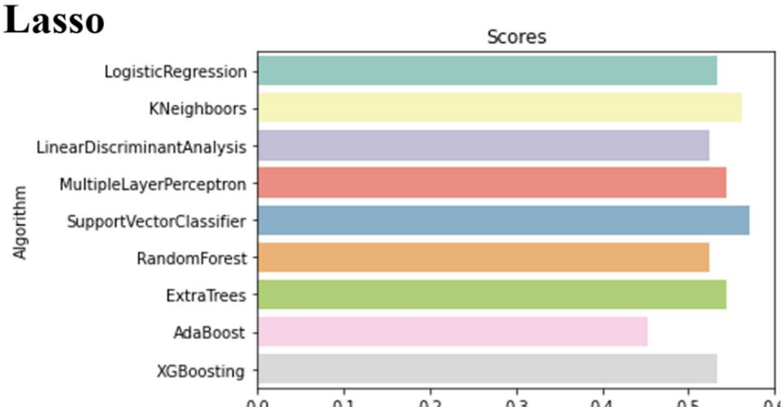


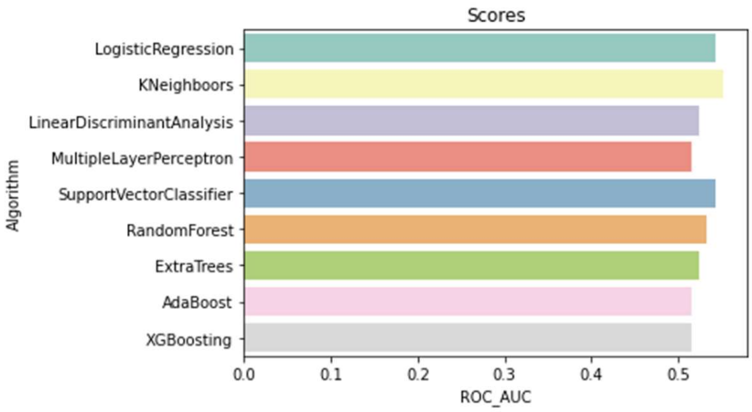
Figure A6. Cross-validation result in dropout-risk group classification using Lasso/ElasticNet features

Lasso



	ROC_AUC	Algorithm
0	0.533693	LogisticRegression
1	0.561995	KNeighboors
2	0.524259	LinearDiscriminantAnalysis
3	0.543127	MultipleLayerPerceptron
4	0.571429	SupportVectorClassifier
5	0.524259	RandomForest
6	0.543127	ExtraTrees
7	0.452830	AdaBoost
8	0.533693	XGBoosting

ElasticNet



	ROC_AUC	Algorithm
0	0.543127	LogisticRegression
1	0.552561	KNeighboors
2	0.524259	LinearDiscriminantAnalysis
3	0.514825	MultipleLayerPerceptron
4	0.543127	SupportVectorClassifier
5	0.533693	RandomForest
6	0.524259	ExtraTrees
7	0.514825	AdaBoost
8	0.514825	XGBoosting

Figure A7. Feature importance result in dropout-risk group classification using SGL features